

sup



(NASA-CR-138800) DERIVATIVE INFORMATION
RECOVERY BY A SELECTIVE INTEGRATION
TECHNIQUE Interim Report (Utah Univ.)

116 p HC \$9.00

CSSL C9B

G3/08

Unclas
43100

N78-27667

DEPARTMENT OF ELECTRICAL ENGINEERING
UNIVERSITY OF UTAH
SALT LAKE CITY, UTAH



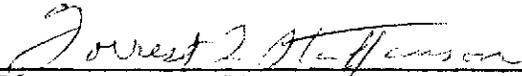
DERIVATIVE INFORMATION RECOVERY BY A
SELECTIVE INTEGRATION TECHNIQUE

Interim Report under
NASA Research Grant NGL 45-003-025

May 1974

by

Marc A. Johnson


Forrest L. Staffanson
Principal Investigator

University of Utah
Electrical Engineering Department

ABSTRACT

A nonlinear stationary homogeneous digital filter DIRSIT (Derivative Information Recovery by a Selective Integration Technique) is investigated. The spectrum of a quasi-linear discrete describing function (DDF) to DIRSIT is obtained by a digital measuring scheme. A finite impulse response (FIR) approximation to the quasi-linearization is then obtained. Finally, DIRSIT is compared with its quasi-linear approximation and with a standard digital differentiating technique. Results indicate the effects of DIRSIT on a wide variety of practical signals.

TABLE OF CONTENTS

ABSTRACT	ii
LIST OF ILLUSTRATIONS AND TABLES	iv
I. INTRODUCTION	1
II. THE DISCRETE DESCRIBING FUNCTION	6
2.1 The Quasi-Linearization	6
2.2 Sampling the DDF	12
2.3 Approximating the Quasi-Linearization	17
III. THE ANALYSIS OF TWO DISCRETE SYSTEMS	19
3.1 The Analysis of the Linear Example	19
3.2 The Analysis of the Nonlinear Example	27
IV. DIRSIT	33
4.1 The Structure of DIRSIT	33
4.2 The "Frequency Response" of DIRSIT	35
4.3 The "Nonlinear Behavior" of DIRSIT	40
4.4 Comparison of DIRSIT to Other Systems	72
4.5 Summary	84
APPENDIX I. A VECTOR SPACE FOR DISCRETE SYSTEM ANALYSIS	99
APPENDIX II. THE 5-T'S METHOD	103
BIBLIOGRAPHY	106

LIST OF ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
3.1a	The measured DDF of the averager. The magnitude, shown here, is similar to $\sin(x)/x$	21
3.1b	The measured DDF of the averager. The phase, shown here, is linear	22
3.2a	The error functions for the DDF of the averager. Part (a) shows the absolute error	23
3.2b	The error functions for the DDF of the averager. The ERF is shown in part (b)	24
3.3a	The measured DDIR of the averager. The real part is shown above	25
3.3b	The measured DDIR of the averager. The imaginary part, shown here, is negligible	26
3.4a	The DDF of LITTLE DIRSIT. The magnitude is displayed here	28
3.4b	The DDF of LITTLE DIRSIT. The magnitude is displayed here	29
3.5	The ERF for the DDF of LITTLE DIRSIT. Note the maximum scale is set at 2.0	30
3.6	The DDIR for LITTLE DIRSIT	31
4.1a	The measured DDF for DIRSIT. The magnitude (linear scale) is shown here	36
4.1b	The measured DDF for DIRSIT. The magnitude (log scale) is shown here	37
4.1c	The measured DDF for DIRSIT. The phase is shown here	38
4.2	The ERF for the DDF of DIRSIT. Note the maximum ERF is set at 2.0 to show the relevant data	39
4.3	The IDFT of the measured DDF of DIRSIT. Only the real part is shown, since the imaginary part is negligible	41

<u>Figure</u>		<u>Page</u>
4.4	An approximated DDIR for DIRSIT	42
4.5a	The approximate DDF to DIRSIT. Part (a) shows the magnitude (linear scale)	44
4.5b	The approximate DDF to DIRSIT. Part (b) shows the magnitude (log scale)	45
4.5c	The approximate DDF to DIRSIT. Part (c) shows the phase	46
4.6	The "first input". A windowed sinusoid with a frequency of 0.1 radians per cycle. This is the first of three inputs run through DIRSIT	47
4.7	The magnitude (log scale) of the first input shown in Fig. 4.6	48
4.8a	The response from DIRSIT to the first input. This is an estimate of the input signal (the first input)	49
4.8b	The response from DIRSIT to the first input. This is an estimate of the first derivative of the input signal (the first input)	50
4.8c	The response from DIRSIT to the first input. This is an estimate of the second derivative of the input signal (the first input)	51
4.9	The log magnitude of the response from the quasi- linear approximation to the first input (the fre- quency description)	52
4.10	The response from the quasi-linear approximation to the first input (the time description). Only the real part is shown, since the imaginary part is negligible	53
4.11	The difference between the response from DIRSIT and the response from the quasi-linear approxima- tion to the first input	55
4.12	The "second input". This is a version of the signal (the first input) contaminated with gaussian noise	56

<u>Figure</u>		<u>Page</u>
4.13	The magnitude (log scale) of the second input shown in Fig. 4.12	57
4.14a	The response from DIRSIT to the second input. This is an estimate of the input signal (the first input)	58
4.14b	The response from DIRSIT to the second input. This is an estimate of the first derivative of the input signal (the first input)	59
4.14c	The response from DIRSIT to the second input. This is an estimate of the second derivative of the input signal (the first input)	60
4.15	The log magnitude of the response from the quasi-linear approximation to the second input (the frequency description)	61
4.16	The response from the quasi-linear approximation to the second input (the time description). Only the real part is shown, since the imaginary part is negligible	62
4.17	The difference between the response from DIRSIT and the response from the quasi-linear approximation to the second input	63
4.18	The "third input". A version of the noisy input signal (the second input) contaminated again with spikes	65
4.19	The magnitude (log scale) of the third input shown in Fig. 4.18	66
4.20a	The response from DIRSIT to the third input. This is an estimate of the first derivative of the input signal (the first input)	67
4.20b	The response from DIRSIT to the third input. This is an estimate of the first derivative of the input signal (the first input)	68
4.20c	The response from DIRSIT to the third input. This is an estimate of the second derivative of the input signal (the first input)	69

<u>Figure</u>		<u>Page</u>
4.21	The log magnitude of the response from the quasi-linear approximation to the third input (the frequency description)	70
4.22	The response from the quasi-linear approximation to the third input (the time description). Only the real part is shown, since the imaginary part is negligible	71
4.23	The difference between the response from DIRSIT and the response from the quasi-linear approximation to the third input	73
4.24a	The response from QLS to the first input. This is an estimate of the input signal (the first input)	75
4.24b	The response from QLS to the first input. This is an estimate of the first derivative of the input signal (the first input)	76
4.24c	The response from QLS to the first input. This is an estimate of the second derivative of the input signal (the first input)	77
4.25a	The response from QLS to the second input. This is an estimate of the input signal (the first input)	78
4.25b	The response from QLS to the second input. This is an estimate of the first derivative of the input signal (the first input)	79
4.25c	The response from QLS to the second input. This is an estimate of the second derivative of the input signal (the first input)	80
4.26a	The response from QLS to the third input. This is an estimate of the first derivative of the input signal (the first input)	81
4.26b	The response from QLS to the third input. This is an estimate of the first derivative of the input signal (the first input)	82

<u>Figure</u>		<u>Page</u>
4.26c	The response from QLS to the third input. This is an estimate of the second derivative of the input signal (the first input)	83
4.27a	The response from QLSDST to the first input. This is an estimate of the input signal (the first input)	85
4.27b	The response from QLSDST to the first input. This is an estimate of the first derivative of the input signal (the first input)	86
4.27c	The response from QLSDST to the first input. This is an estimate of the second derivative of the input signal (the first input)	87
4.28a	The response from QLSDST to the second input. This is an estimate of the input signal (the first input)	88
4.28b	The response from QLSDST to the second input. This is an estimate of the first derivative of the input signal (the first input)	89
4.28c	The response from QLSDST to the second input. This is an estimate of the second derivative of the input signal (the first input)	90
4.29a	The response from QLSDST to the third input. This is an estimate of the first derivative of the input signal (the first input)	91
4.29b	The response from QLSDST to the third input. This is an estimate of the first derivative of the input signal (the first input)	92
4.29c	The response from QLSDST to the third input. This is an estimate of the second derivative of the input signal (the first input)	93
4.30a	The measured DDF for the modified DIRSIT. The magnitude (linear scale) is shown above	94
4.30b	The measured DDF for the modified DIRSIT. The magnitude (log scale) is shown above	95

<u>Figure</u>		<u>Page</u>
4.30c	The measured DDF for the modified DIRSIT. The phase is displayed here	96
4.31	The ERF for the DDF of the modified DIRSIT	98

I. INTRODUCTION

This research is intended to supply information on the effects of DIRSIT (Derivative Information Recovery by a Selective Integration Technique) on practical signals. This information is obtained by developing a discrete describing function (DDF).

DIRSIT is a nonlinear homogeneous stationary digital filter used to reduce data contaminated by gaussian noise and shot noise [4] [5]. The early investigators found that DIRSIT has almost complete immunity to noise spikes and can track the second integral of a step function with very little error. Unfortunately, little is known of the effect of this filter on practical signals. What information is available is limited to specific input functions, and is qualitative and empirical [4], [5].

The describing function is a traditional method to "describe" nonlinear systems. The describing function is the transfer function of a quasi-linearization of the nonlinear system [1]. The quasi-linearization has the property that its output is as close as possible (in some defined sense) to the output of the nonlinear system for some specific set of input functions.

The describing function method is useful as a description provided the nonlinear system is "almost linear" in the region of interest. A system is "almost linear" on a set of input functions (denoted 'S'), provided that the system almost obeys linear superposition over S; that is, the output of the system for an input consisting of any finite sum of functions from S, is "close" to the sum of

the outputs for each individual input. Traditionally, two functions are "close" if the root mean square of their difference is small. This is the topology used throughout this paper. How close the system must come to obeying linear superposition depends upon the application.

The sinusoidal describing function is a describing function where the set S consists of sinusoids. The output of the sinusoidal quasi-linear approximation is the closest linear approximation to the output of the nonlinear system. Thus, by Fourier theory, the quasi-linear output is the fundamental component of the nonlinear output [1]. The describing function of a linear system, then, is the transfer function of that linear system. The sinusoidal describing function began as a heuristic extension of the linear system transfer function concept. It is an input amplitude dependent version of a system transfer function. Measurement of the describing function is feasible only if the system is homogeneous (amplitude independent).

The discrete describing function (DDF) developed here is analogous to the continuous sinusoidal describing function. As in the continuous case, the DDF is a quasi-linear approximation to the given system. The output of the quasi-linearization for a sampled sinusoidal input is the "fundamental component" of the system output. The quasi-linear approximation minimizes the mean-square difference between the actual and the approximate output for a sampled sinusoidal input.

The concept of a describing function to "describe" a discrete system is not new. In 1959, Ben C. Kuo introduced what he called the "Z-transform-describing function" [2]. The Z-transform-describing function for a discrete nonlinear element is defined as the ratio of the Z-transform of the output from the element, to the Z-transform of the input to the element. The input to the nonlinear element is restricted to a specific function; in most cases, that function is a sampled sinusoid. The Z-transform-describing function may be used in the standard linear analysis technique to replace the nonlinear element with no approximation errors, provided that the exact input to the nonlinear element is known a priori. Dr. Kuo was able to utilize this technique to analyze the behavior of a feed-back loop containing a relay.

The discrete describing function (DDF) developed here may also be used in evaluating discrete nonlinear feedback systems. The procedure would be to substitute the DDF for the nonlinear element in the standard linear analysis technique. An advantage in using the DDF is that the exact input need not be known. A single analysis is valid for a rather wide range of inputs to the nonlinear element. A disadvantage in using the DDF is that there is an approximation being made. This problem may be circumvented by using a "remnant" generator in much the same way as is done in the continuous case, but this would then reduce to the Z-transform-describing function approach.

The use of the DDF to evaluate discrete systems is similar to the use of the sinusoidal describing function to evaluate continuous

systems. The evaluation of a nonlinear system in this context is a frequency description of the input output relation of the system. This description is necessarily limited, and valid only on the set S where the system is almost linear. The use of the DDF in this paper will be limited to this type of evaluation of discrete systems.

A vector space approach is used to sample the DDF. The vector space approach utilizes the projection theorem [3]. This is possible since the quasi-linear output is the "best" (closest) linear approximation to the nonlinear system. An algorithm is developed to obtain samples of the frequency response of a quasi-linear approximation to any homogeneous (or almost homogeneous) discrete system. These samples are then used to realize an approximation to the quasi-linearization by the standard 5-T's method [6]. The projection theorem approach also enables the calculation of how close the quasi-linearization is to the system under investigation. Thus, the method provides a means of evaluating the results.

Two simple systems, one linear and one nonlinear, are evaluated using this technique. The DDF for the linear system turns out to be identical to the linear system transfer function as expected. The nonlinear system is chosen because it is simple and has an obvious similarity to DIRSIT.

Finally, the DDF of DIRSIT is obtained. DIRSIT is then compared with its quasi-linear approximation and with a standard technique of differentiating and smoothing digital data. This results in

a greater understanding of the linear and nonlinear effects of DIRSIT on a large class of practical signals.

II. THE DISCRETE DESCRIBING FUNCTION

The discrete describing function (DDF) developed here is analogous to the continuous sinusoidal describing function. The use of the DDF in this paper is limited to that of "evaluating" discrete nonlinear homogeneous (or almost homogeneous) systems.

The evaluation of a discrete system by its DDF is similar to the evaluation of a continuous system by its sinusoidal describing function. The most important similarity is that the quasi-linearization is the closest linear approximation to the nonlinear system. For the continuous case, this means extracting the fundamental component from the nonlinear output. This is a consequence of Fourier theory. For the discrete case, however, the situation is not so simple. The Fourier theory does not apply. The projection theorem is utilized in order to provide the existence and uniqueness of the approximation.

2.1 The Quasi-Linearization

The quasi-linear approximation to a system is a linear system. The linear system is referred to as the quasi-linear system, or quasi-linearization (of the approximated system). The impulse response of the quasi-linearization is referred to here as the discrete describing impulse response (DDIR). The Z-transform of the DDIR evaluated on the unit circle in the complex plane is the DDF of the approximated discrete system. The DDF for a discrete system is a complex-valued function of the real variable w (in units of radians per sample) over the

interval $(-\pi, \pi]$. (Recall that the Nyquist frequency is π radians per sample.) Thus, if $h(n)$ is the impulse response of a quasi-linear approximation to a nonlinear system, then $h(n)$ is the DDIR of that system. If $H'(z)$ is the Z-transform of $h(n)$, then $H(w) = H'[z = \exp(jw)]$ is the DDF of the nonlinear system. The expression $H'[z = \exp(jw)]$ refers to the Z-transform of the system impulse response evaluated on the unit circle in the complex plane at $z = \exp(jw)$.

The quasi-linearization minimizes the mean-square difference between the actual and the approximate output for a sampled sinusoidal input. The output of the quasi-linearization of a system is the closest linear approximation to the output of the system. The linear approximation is based on the sampled complex exponentials of the fundamental frequency, or equivalently, the sampled sine and cosine functions of the fundamental frequency.

This may be seen as follows: Let the input to the system and the output from the system be of the form,

$$f(n) = \cos nw \tag{2.1}$$

$$g(n) = a \exp(jnw) + b \exp(-jnw) + e(n) \tag{2.2}$$

respectively, where n ranges over all integers, w ranges over $[0, \pi]$, j is the square root of -1 , and a and b are complex constants. The restriction of the range of w is expedient since, by the sampling theorem [6], if w is greater than π , aliasing will occur.

The complex numbers a and b are obtained by projecting $g(n)$ on $\exp(jn\omega)$ and $\exp(-jn\omega)$, respectively, so that the expression

$$g'(n) = a \exp(jn\omega) + b \exp(-jn\omega) \quad (2.3)$$

represents the closest linear approximation to $g(n)$ based upon $\exp(jn\omega)$ and $\exp(-jn\omega)$. These basis functions form an orthonormal set (ONS). They are chosen because the sampled complex exponential is an eigenfunction for all linear discrete systems. Since the input (Eq. 2.1) is a linear combination of the basis functions, the response to $f(n)$ in Eq. 2.1 from a linear discrete system must also be some linear combination of the basis functions. In other words, for an input of the form (Eq. 2.1), this basis set spans the space of all possible outputs from any linear system. Thus, $g'(n)$ is the "best" linear approximation to $g(n)$.

By using Eq. 2.3, Eq. 2.2 may be rewritten in the form

$$g(n) = g'(n) + e(n) \quad (2.4)$$

The function $e(n)$ is orthogonal to $g'(n)$ and therefore contains the "nonlinear part" of $g(n)$. The magnitude of $e(n)$ is a quantitative measure of the size of this "nonlinear part". The ratio of the magnitude of $e(n)$ to the magnitude of $g'(n)$, a function of frequency, is defined here as the error ratio function (ERF). The ERF is a measure of the relative portion of the nonlinear part of $g(n)$. If

the value of ERF is much less than one, then the system exhibits a small nonlinear effect on its input. If the value of ERF is not small, then the system exhibits a large nonlinear effect on its sampled sinusoidal input. In many applications, the ERF could conceivably be utilized to determine whether or not the DDF is an adequate "description" of a nonlinear element.

If a discrete homogeneous stationary system is "almost linear" on some region of sampled sinusoids, then the ERF on that region is small. This may be seen as follows: Let the input to the system be

$$f(m) = \exp(jmw) \quad (2.5)$$

and denote the corresponding output by

$$g(m) = h(w,m) \quad (2.6)$$

Since the system is stationary, it follows that the response to

$$f(m) = \exp[j(m+n)w] \quad (2.7)$$

is

$$g(m) = h(w,m+n) \quad (2.8)$$

for all n . Since the system is homogeneous, it follows that the response to

$$f(m) = \exp(jnw) \exp(jmw) \quad (2.9)$$

is

$$g(m) = \exp(jnw) h(w,m) \quad (2.10)$$

But the inputs of Eqs. 2.7 and 2.9 are identical, so that the corresponding outputs of Eqs. 2.8 and 2.10 must also be identical. For the particular case where $m = 0$, this latter identity becomes

$$h(w,n) = h(w,0) \exp(jnw) \quad (2.11)$$

Thus the response to Eq. 2.5 may be written

$$g(m) = h(w,0) \exp(jmw) \quad (2.12)$$

Now, suppose the system is "almost linear" at w , that is, with respect to $\exp(jnw)$ and $\exp(-jnw)$. It follows that the response to the input function of Eq. 2.1 is given by

$$g(n) = [h(w,0) \exp(jnw) - h(-w,0) \exp(-jnw)]/2 + e(n) \quad (2.13)$$

where $e(n)$ is small. Thus, the desired conclusion is established!

Wherever the system is almost linear, its ERF is small. This result can easily be extended to "almost homogeneous" systems.

It must be remembered that the converse of the above result is not necessarily true. If a nonlinear system exhibits small nonlinear effects over a range of input frequencies, it does not follow conclusively that the system is "almost linear" for these frequencies. This

is due to the fact that without further assumptions about a nonlinear system, the individual outputs of particular functions are, in general, no indication of what the outputs of their sum will be. Thus, the ERF is a quantitative measure of how nonlinear a system is, i.e., where the quasi-linear approximation is a good one, and also indicates the frequency range where the system may be "almost linear".

In many practical cases in general, and in the systems examined in this paper in particular, it is assumed that the ERF is an adequate indication of where the system is almost linear. In general, it seems that it is possible to determine where a system is almost linear only by building the set of candidates by trial and error. However, where a homogeneous system exhibits a nonlinearity that effects primarily frequencies well outside the region of interest, it seems reasonable to assume the ERF is an indication of where the system is almost linear. It is well to remember, however, that what frequencies the "non-linear character" of the system effects is a matter of experience and judgment of the user.

Use will be made of two more results about the quasi-linearization. First, the quasi-linearization exists; and second, the DDIR is real. These results are assumed in most continuous discussions about describing functions. Thus, since they are probably somewhat obvious, only a rather brief argument for them will be given here.

The existence of the quasi-linearization follows from the existence part of the projection theorem. Whenever the input to a discrete

system is of the form of Eq. 2.1, and the output is periodic, then the output can be put in the form of Eq. 2.2. This is a direct consequence of the fact that the ONS basis used to generate $g'(n)$ in Eq. 2.3 is finite, and therefore generates a closed subspace.

The quasi-linear response to an input of the form of Eq. 2.1 may always be put in the form of Eq. 2.3. But a and b are the values of the DDF at w and $-w$, respectively (since the representation of $g'(n)$ is unique!). It will be shown in the next section that $b = a^*$. Thus, the DDF exhibits an even real part and an odd imaginary part. But this is equivalent to forcing the DDIR to be real! Note that the fact that $g(n)$ can be put in the form of Eq. 2.2 again follows from the projection theorem.

The projection theorem, then, provides the means to obtain the unique "best" linear approximation to a system with a sampled sinusoidal input. This "best" linear approximation is the quasi-linear approximation. The projection theorem also provides a quantitative method to evaluate this approximation.

2.2 Sampling the DDF

The value of the DDF for a specific frequency may be obtained by projecting the sampled sinusoidal response of that system on the corresponding sampled complex exponential. The coefficient of the projection, a complex number, is half the value of the DDF at that frequency. For example, if the system has the input given by Eq. 2.1, and an output given by Eq. 2.2, then the value of the DDF of that system

at w is $2a$. This may be demonstrated by explicitly solving for a and b in Eq. 2.2, and relating the results to the known behavior of discrete linear systems.

The solution of a and b in Eq. 2.2 is given by

$$\begin{aligned}
 a &= (g(n), \exp(jnw)) \\
 b &= (g(n), \exp(-jnw)) \\
 &= [(g(n), \exp(jnw))]^* \\
 &= a^*
 \end{aligned} \tag{2.14}$$

wherein '*' indicates the complex conjugate. A short account of the vector space used in this paper and its associated inner product is given in Appendix I. Using the results of Eq. 2.14, Eq. 2.3 may be written in the form

$$g'(n) = \text{Re}[2 a \exp(jnw)] \tag{2.15}$$

Any linear discrete system with an input of the form

$$x(n) = \exp(jnw) \tag{2.16}$$

has an output of the form

$$y(n) = H'[z = \exp(jw)] \exp(jnw) \tag{2.17}$$

In Eq. 2.17, H' denotes the Z-transform of the impulse response of the system. If the input to the system is given by

$$x(n) = \exp(-jnw) \quad (2.18)$$

then the output is given by

$$y(n) = H'[z = \exp(-jw)] \exp(-jnw) \quad (2.19)$$

Thus, since the input function of Eq. 2.1 is a linear combination of Eqs. 2.16 and 2.18, the output must be the same linear combination, so that the response to an input of the form of Eq. 2.1 is given by half the sum of the outputs of Eqs. 2.17 and 2.19. Since the discrete homogeneous linear system under consideration is assumed to have a real impulse response (which implies a Z-transform that exhibits complex conjugate symmetry), the output from the system for the input of Eq. 2.1 can be written

$$y(n) = 2\text{Re}(H'[z = \exp(jw)] \exp(jnw)) \quad (2.20)$$

Utilizing the notation $H(w) = H'[z = \exp(jw)]$, Eq. 2.20 may be written in the form

$$y(n) = \text{Re}[H(w) \exp(jnw)] \quad (2.21)$$

This result in connection with Eq. 2.15 yields the conclusion

that

$$\text{DDF}(w) = 2 a \quad (2.22)$$

for w in $(0, \pi)$. Since the DDIR is real, its Z-transform has complex conjugate symmetry, so that the values of $\text{DDF}(w)$ for negative w can be obtained from knowing $\text{DDF}(w)$ for positive w . In particular, the magnitude of the DDF is even and the angle (phase) is odd.

Combining Eqs. 2.22 and 2.14 yields the results

$$\text{DDF}(w) = 2(g(n), \exp(jnw)) \quad (2.23)$$

for w in the open region $(0, \pi)$. For the special case where $w = 0$ and $w = \pi$, the input function of Eq. 2.1 is a constant. The ONS basis consists, under these circumstances, of a single constant, so that the value of the DDF for these cases is given by

$$\text{DDF}(w) = (g(n), 1) \quad (2.24)$$

The two expressions of Eqs. 2.23 and 2.24, together with the symmetry relations, theoretically give a complete recipe for the evaluation of the DDF.

Applying these expressions, however, presents immediate practical difficulties. The inner product in Eq. 2.23 can never be calculated in practice since it involves knowing $g(n)$ for all integers n . A windowing procedure of some sort must be used on the terms involved in Eq. 2.23 to approximate the solution. There are two cases to

consider: when $2\pi/w$ is rational and when it is not.

Suppose that $2\pi/w$ is rational, i.e., there exists two integers, say N and K , such that $Nw = 2\pi K$. Then, if the input is of the form

$$x(n) = \cos[(n - k)w] \quad (2.25)$$

for some integer k , it follows that $x(n + N) = \cos[(n + N + k)w] = \cos[(n + k)w + Nw] = \cos[(n + k)w] = x(n)$, i.e., $x(n)$ is periodic of period N . In this case, the inner product of Eq. 2.23 reduces to a finite sum. This assumes the range of the summation covers a "representative region". (See Appendix I.)

The other case to consider is when $2\pi/w$ is irrational. In this case, the input is not periodic. The procedure then is to choose N such that the approximation by truncating the sum will be small. Intuitively, the sequence is "almost periodic" (i.e., $x(n)$ is close to $x(n + N)$ for all n), when there exists two integers N and K such that wN is nearly $2\pi K$. Then the finite sum may be utilized, except that an approximation is being made.

Thus far, then, the frequency response of the quasi-linearization at w radians per sample is obtained (or approximated) by inputting $f(n)$ in Eq. 2.1, $n = 0, 1, \dots, N - 1$ for an N where Nw is exactly (or approximately) a multiple of 2π .

Unfortunately, the solution of one problem presents another. If an input to the system under study is to be finite, the system, being in general nonlinear, will exhibit transient response. This

response is not to be considered part of its legitimate output, since for a true sinusoidal input, transient response is nonexistent. The solution could be any of several. The solution chosen here was to input the sequence for a fixed number of inputs, assuming a priori, that would exhaust transients. Note that the derivation of Eqs. 2.15 and 2.21 are valid if n is replaced by $n - k$ throughout, where k is an arbitrary integer, so that Eq. 2.22 is still valid, and Eq. 2.23 takes the form

$$DDF(w) = 2(g(n - k), \exp(jnw - jkw)) \quad (2.26)$$

Equation 2.24 may be written

$$DDF(w) = (g(n - k), 1) \quad (2.27)$$

Thus, the value of the $DDF(w)$ for w in the range between zero and π may be calculated with a finite sum given by Eq. 2.26. This calculation is accurate for frequencies whose ratio to 2π is rational and otherwise an approximation. The negative frequencies may be obtained from the positive ones through symmetry relations.

2.3 Approximating the Quasi-Linearization

The results of following the above recipe will yield, at best, discrete samples of the Z-transform of the impulse response of the quasi-linearization. There remains the problem of converting these samples into usable information. The approach used in this study is

to sample the frequency response of the system at even intervals a power of 2 times. This allows a situation where the ratio of the samples frequency and 2π is always rational and where the periodicity of the sequences is well within the range of the computer memory. Thus, no error need be introduced at that point. This also allows a power of 2 samples of the frequency response of the quasi-linearization to be available. This is exactly the number of samples necessary to utilize the FFT algorithm, which is implied in the 5-T's method of realizing a FIR filter. (See Appendix II.)

The final algorithm to realize an approximation to the quasi-linearization of a discrete system is now complete. Sample the frequency band a power of two times. For each frequency sample, input a fixed number of samples of the sinusoid of that frequency in the system to dissipate any transients. Then input one period of samples. Project the system response on the corresponding sampled complex exponential (via Eqs. 2.26 or 2.27). This coefficient is the value of the DDF at that frequency sample. This result is a power of 2 samples of the frequency response of the quasi-linearization. Next step is to apply the 5-T's technique to realize an approximation to the quasi-linearization. The resulting approximation may then be used to operate on an input signal via the fast convolution techniques [6]. This may then be compared to the response of the system under study in order to determine what effect the nonlinearity has on the input signals. The process also provides a technique for evaluating the accuracy of the results.

III. THE ANALYSIS OF TWO DISCRETE SYSTEMS

Two homogenous discrete digital systems are evaluated to illustrate the use of the DDF technique. The first system is a simple linear system, included to illustrate the fact that the technique will reproduce a linear system. The second system is a nonlinear system. This system is included to illustrate the use of the DDF technique on a nonlinear system.

For each system, the DDF is sampled evenly on the interval $(-\pi, \pi]$. There is a sample at 0 and π , and 255 evenly spaced samples in the open interval from 0 to π . There is, then, a total of 512 samples on $(-\pi, \pi]$. When the frequency response is plotted, only points from 0 to π are plotted, since negative frequencies obey symmetry relations.

3.1 The Analysis of the Linear Example

The linear system is a simple averager. The output at a point k is the average of the k th input point together with the previous 8 input points. The output may be written in the form

$$g(k) = [f(k) + f(k - 1) + \dots + f(k - 8)]/9 \quad (3.1)$$

The impulse response of this system is identically zero except for the zeroth to the eighth point inclusive, where it has the value of $1/9$. The DDF of the impulse response is a $\sin(x)/x$ type curve that has been wrapped around upon itself.

The DDF of a linear system is obtained using the developed

measuring technique. The DDIR of a linear system is the impulse response of that system. The DDF of a linear system is the Z-transform of the system impulse response evaluated on the unit circle in the complex plane. These facts can be seen by inspecting Eqs. 2.2 and 2.3. Recall that the orthonormal basis functions $\exp(jn\omega)$ and $\exp(-jn\omega)$ span the space of possible outputs from a linear system and that $g'(n)$ in Eq. 2.3 was obtained by projecting the system output $g(n)$ in Eq. 2.2 on that space. Thus, for a linear system, $g(n)$ will be in the space and the closest approximation to it will be itself. The ERF is expected to be nearly identically zero.

The results of the measurements made on the linear "averager" are contained in Fig. 3.1 to Fig. 3.3. Fig. 3.1 is the measured DDF of the "averager". The measured DDF (Fig. 3.1) indeed resembles the $\sin(x)/x$ curve. The absolute error function is displayed in Fig. 3.2a. The ERF is displayed in Fig. 3.2b. The spikes occurring approximately at even intervals are due to the fact that the ERF is the ratio of the magnitude of the "nonlinear part" to the magnitude of the "linear part" of the system output, and the magnitude of the "linear part" is nearly zero at these frequencies (see Fig. 3.1), causing machine overflow. The IDFT of the measured DDF was obtained and plotted in Fig. 3.3. Fig. 3.3a shows the impulse response of the linear system exactly to at least three places. Fig. 3.3b is included to show the imaginary part to be null within the order of the errors expected in computing the IDFT. Thus, the measured DDF description is very close to the linear transfer function description.

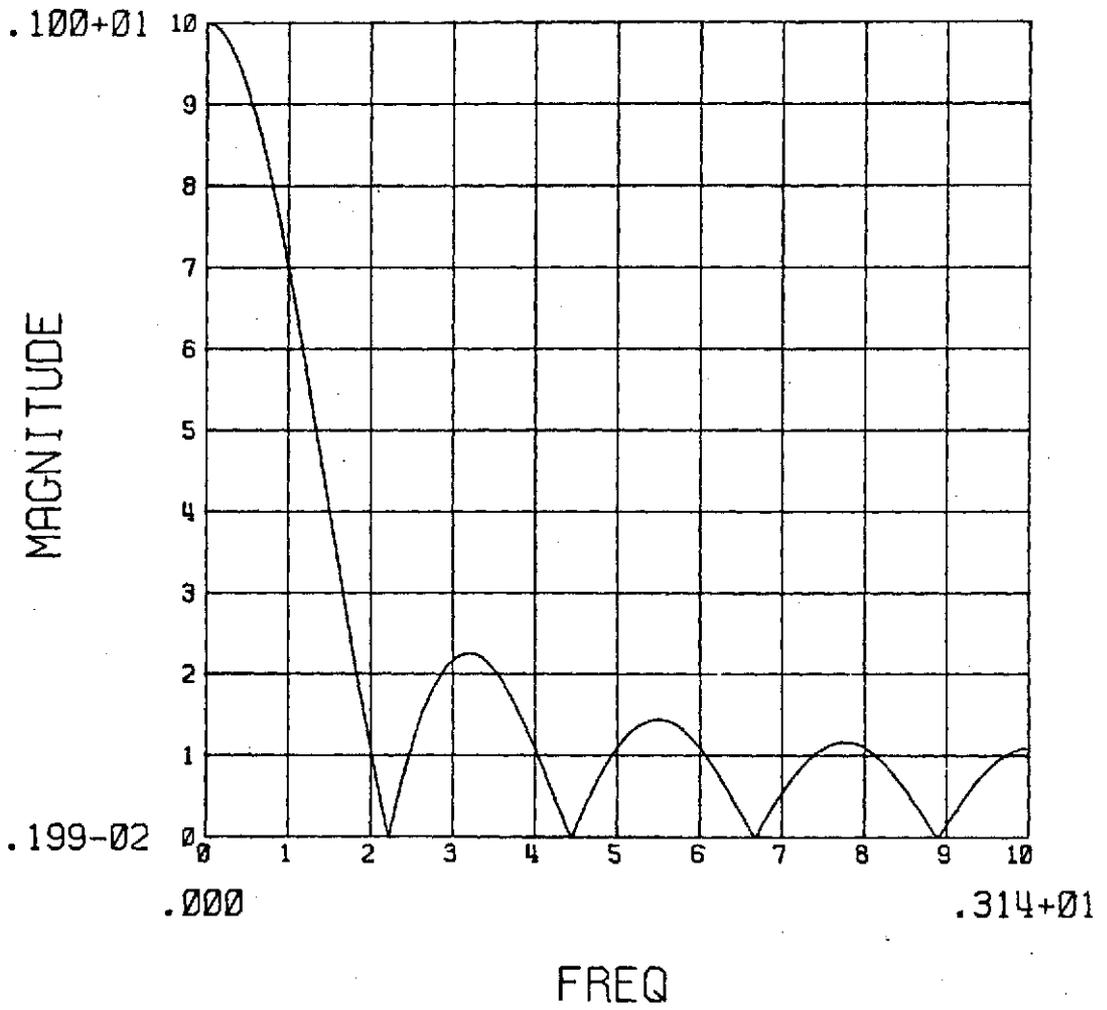


Fig. 3.1a. The measured DDF of the averager. The magnitude, shown here, is similar to $\sin(x)/x$.

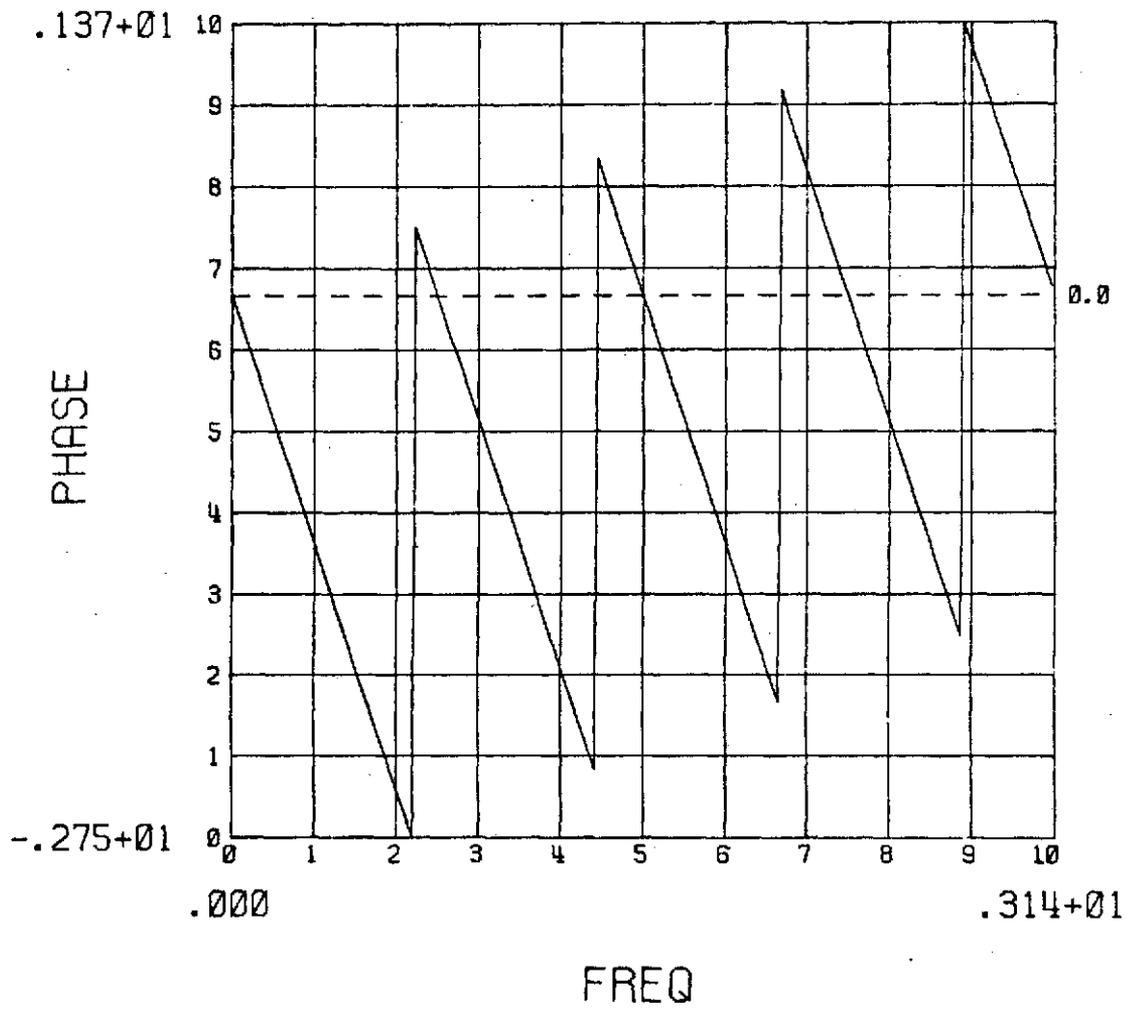


Fig. 3.1b. The measured DDF of the averager.
The phase, shown here, is linear.

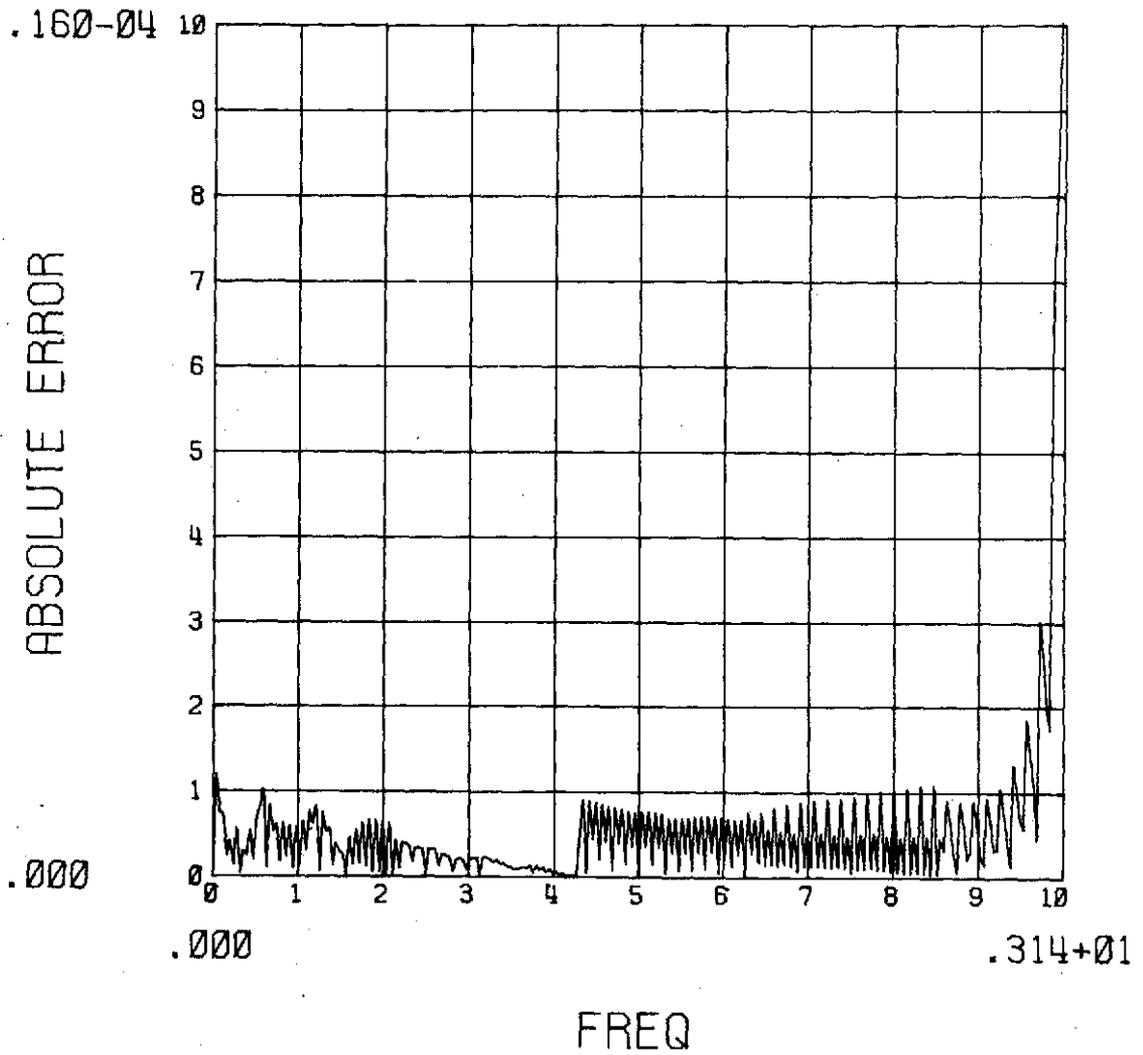


Fig. 3.2a. The error functions for the DDF of the aver-
ager. Part (a) shows the absolute error.

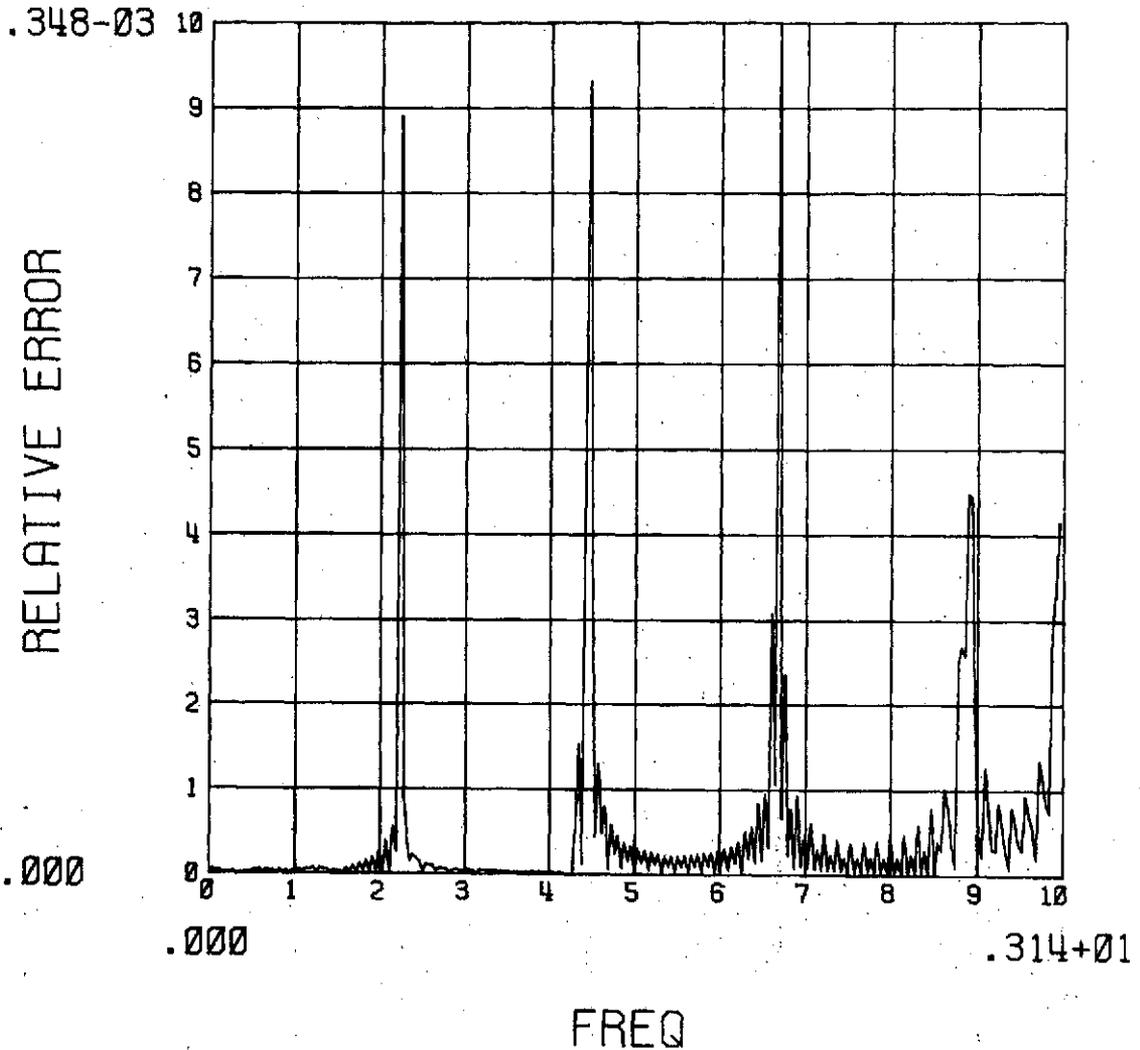


Fig. 3.2b. The error functions for the DDF of the averager. The ERF is shown in part (b).

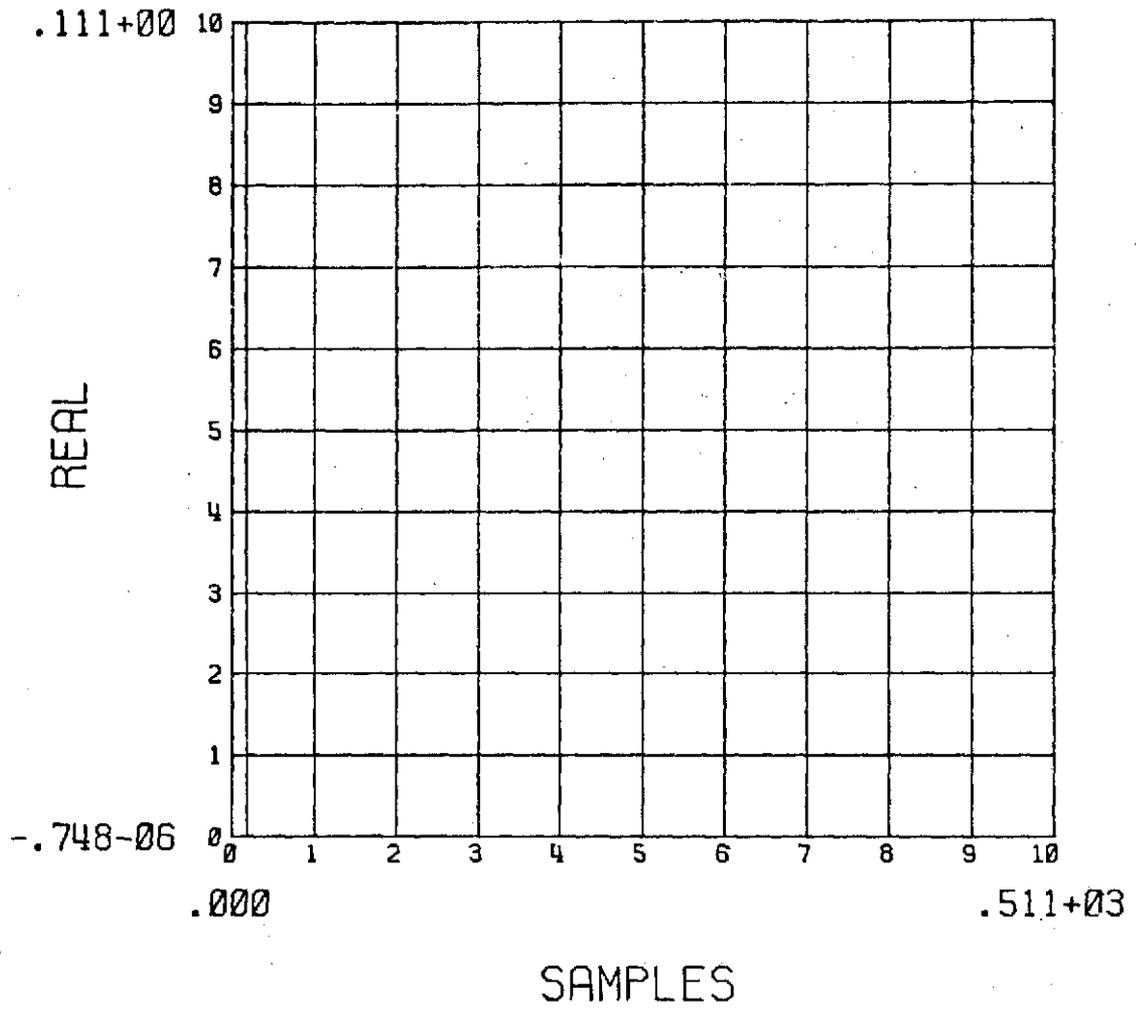


Fig. 3.3a. The measured DDIR of the averager.
The real part is shown above.

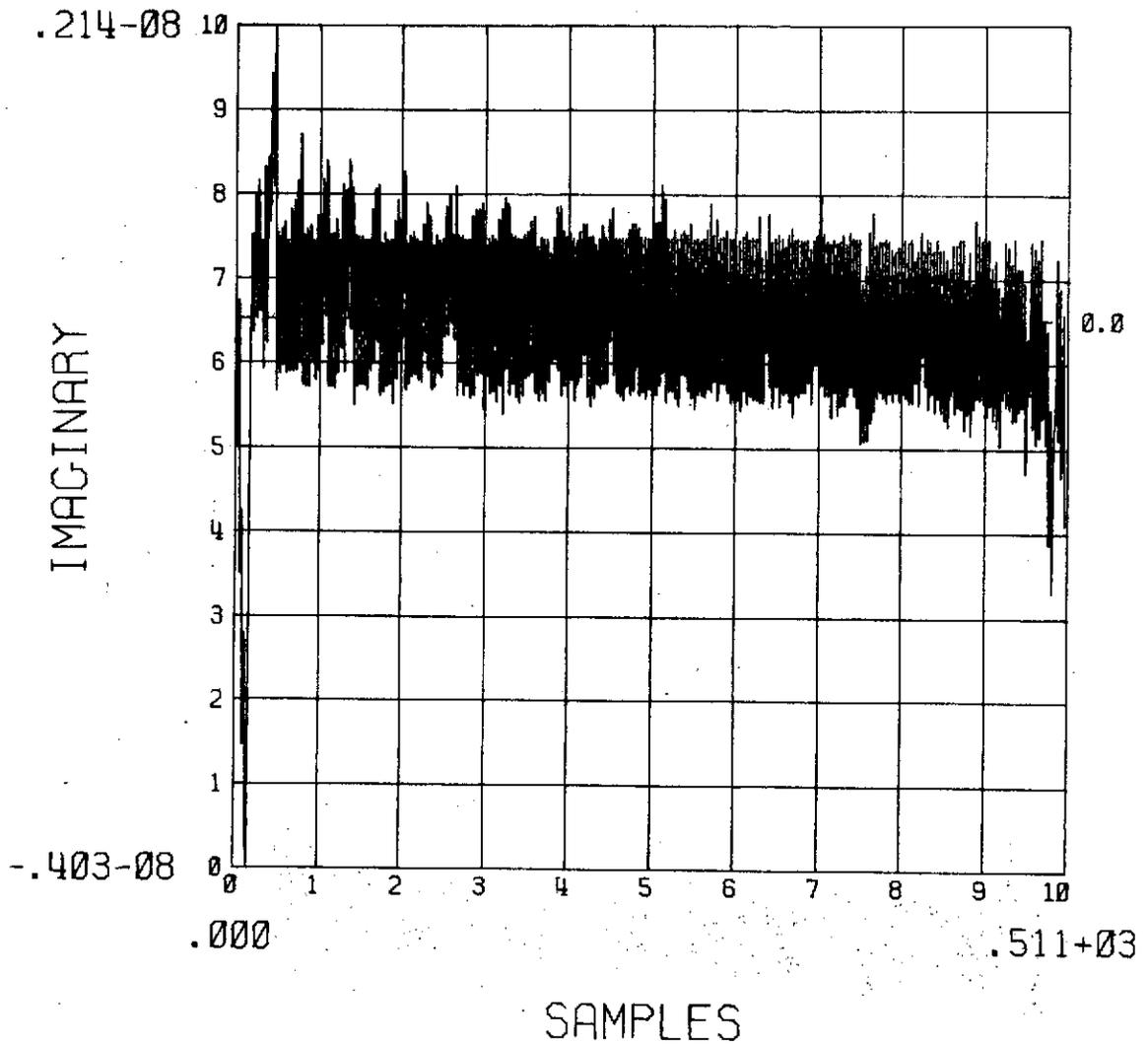


Fig. 3.3b. The measured DDIR of the averager. The imaginary part, shown here, is negligible.

3.2 The Analysis of the Nonlinear Example

The nonlinear system extracts the median of a set of 9 points. For reference purposes, this system will be referred to here as "LITTLE DIRSIT". The output of LITTLE DIRSIT at a point k is the median of the set of points consisting of the current (k th) point together with the previous eight points. LITTLE DIRSIT is a highly discontinuous system, and will pass a square pulse unaffected, provided that its length is more than four points. LITTLE DIRSIT was chosen as a nonlinear example because it has some obvious similarities to DIRSIT.

The DDF of LITTLE DIRSIT was measured and is displayed in Fig. 3.4. The corresponding ERF is plotted in Fig. 3.5. The ERF indicates that the "description" is rather good for frequencies zero to about .6 radians per sample. There are other regions that are good but this region will serve to illustrate the position taken here. Fig. 3.4 indicates that over these frequencies, LITTLE DIRSIT behaves much like the linear "averager", except that it has some "notches" in the frequency band. Fig. 3.6 is a plot of the real part of the DDIR of LITTLE DIRSIT obtained from the measured DDF samples. The imaginary part of the DDIR is negligible. This is a first FIR approximation to the "real" DDIR.

An important property of the DDIR of this nonlinear discrete system is shown in Fig. 3.6. The DDIR is not realizable! The DDIR is not identically zero for negative values of the argument n . (Recall that the IDFT of samples of a frequency response yield a periodic

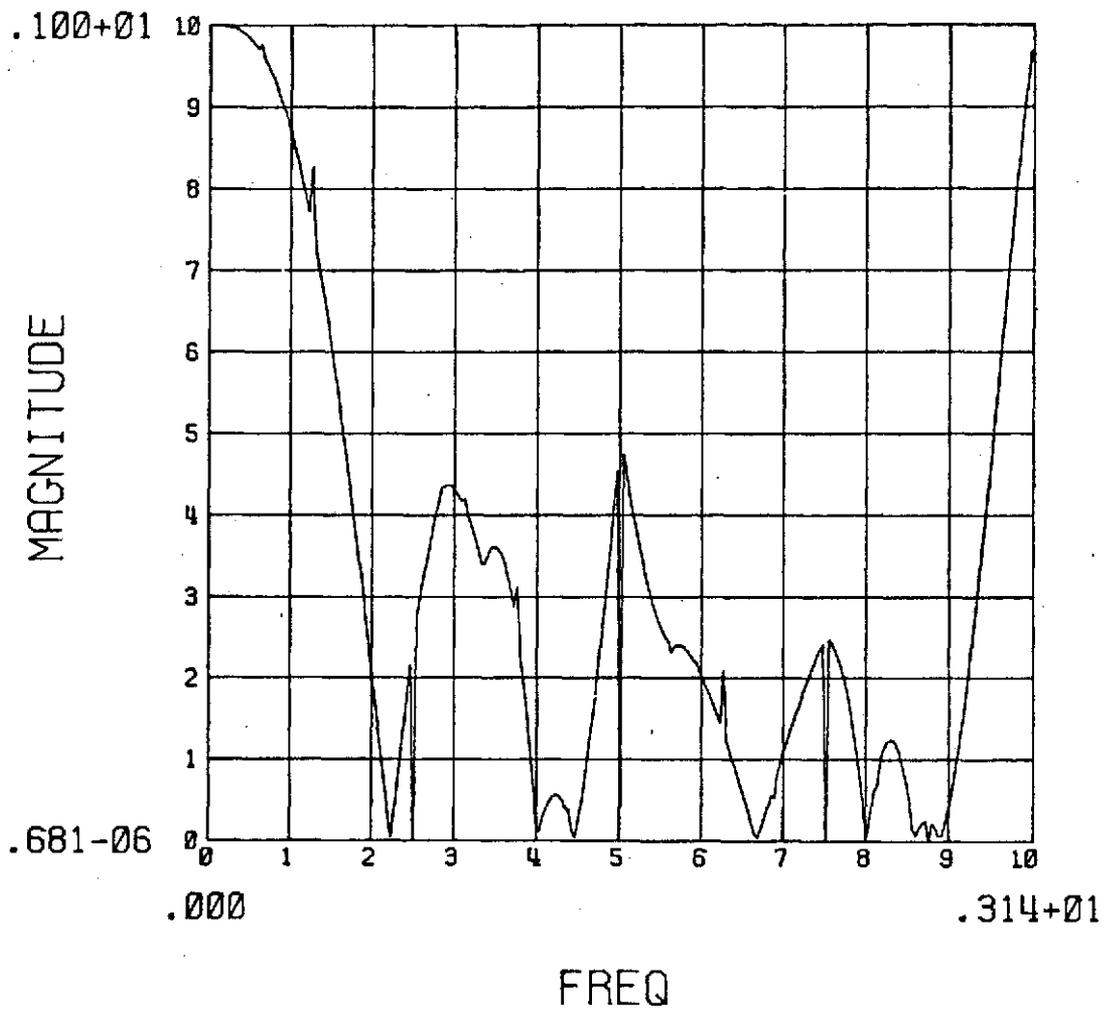


Fig. 3.4a. The DDF of LITTLE DIRSIT. The magnitude is displayed here.

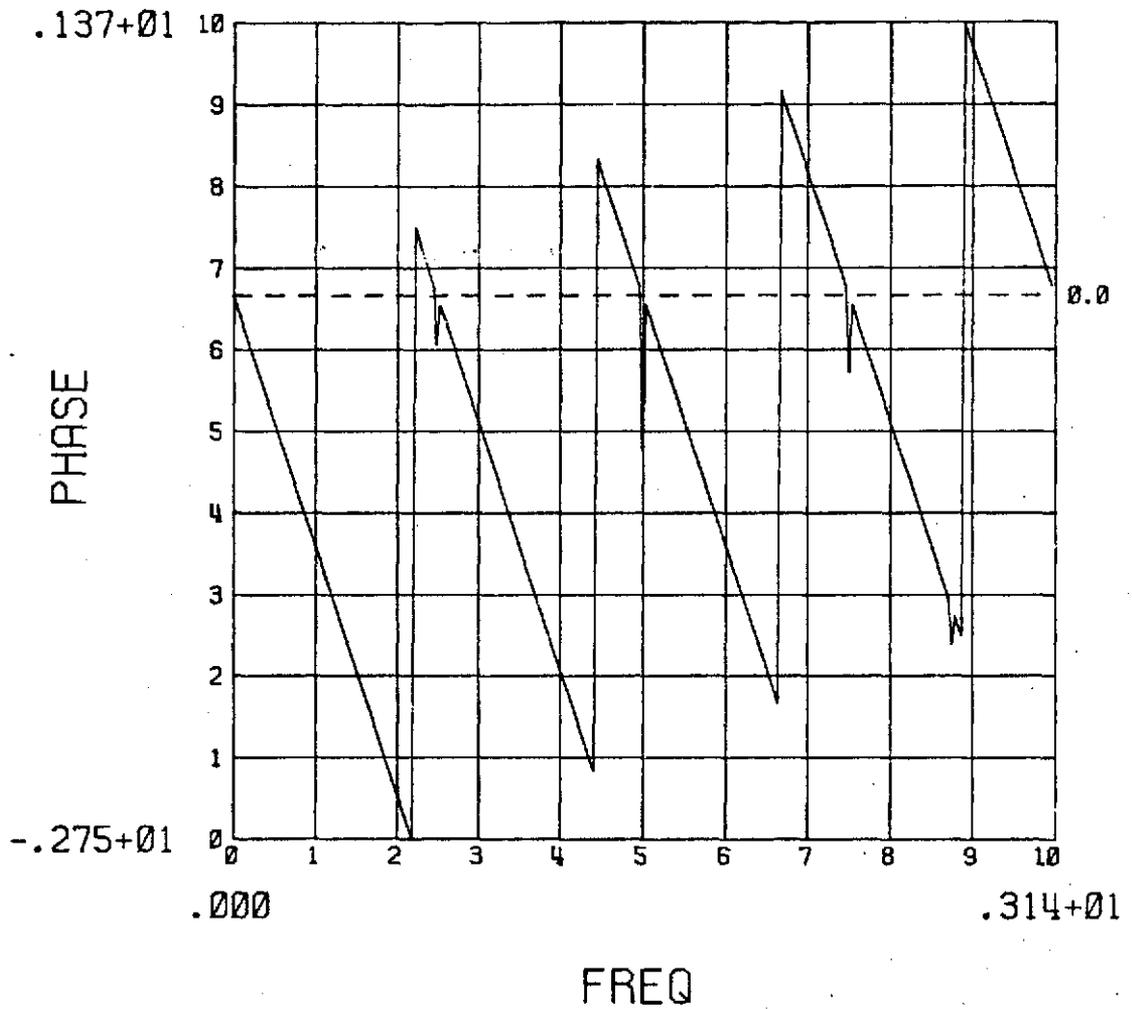


Fig. 3.4b. The DDF of LITTLE DIRSIT. The magnitude is displayed here.

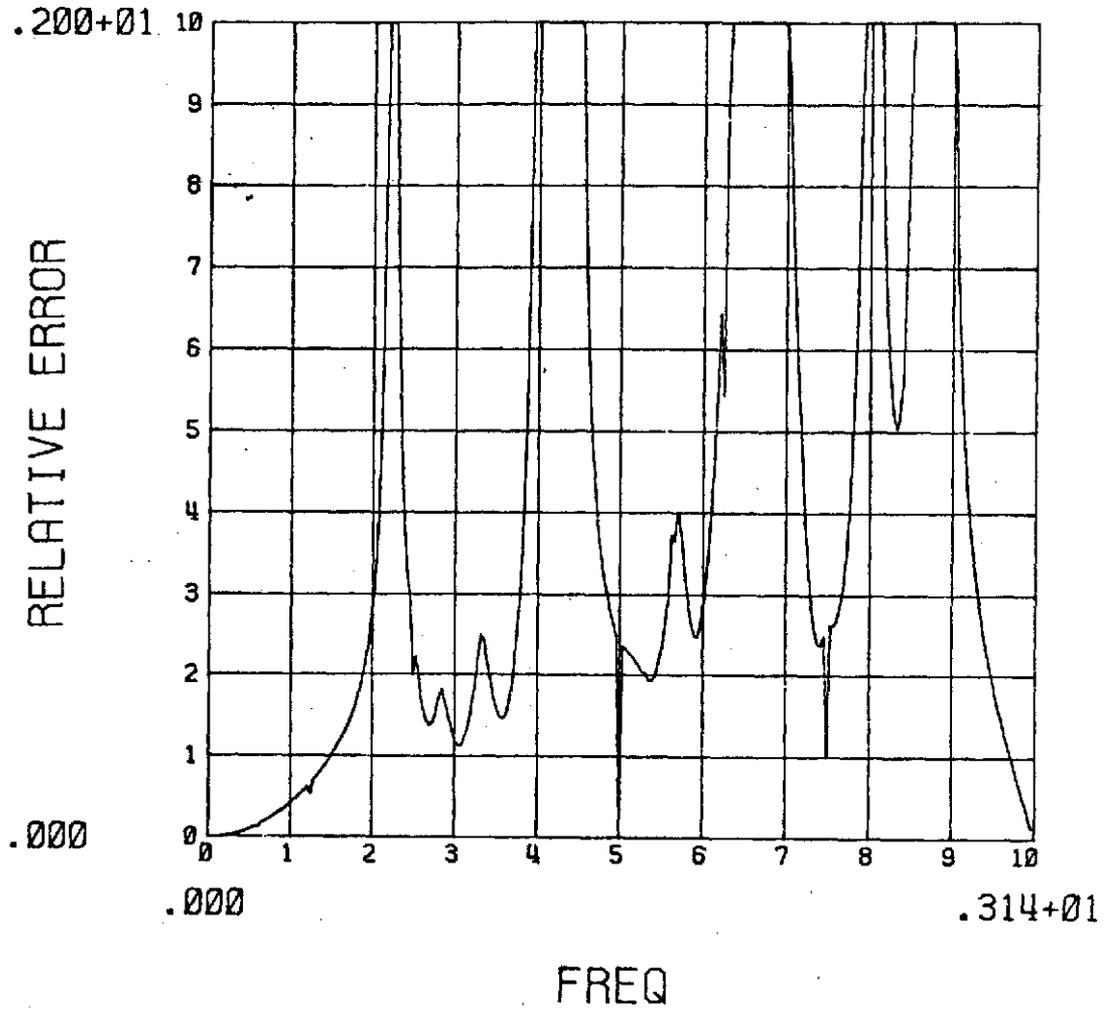


Fig. 3.5. The ERF for the DDF of LITTLE DIRSIT.
 Note the maximum scale is set at 2.0.

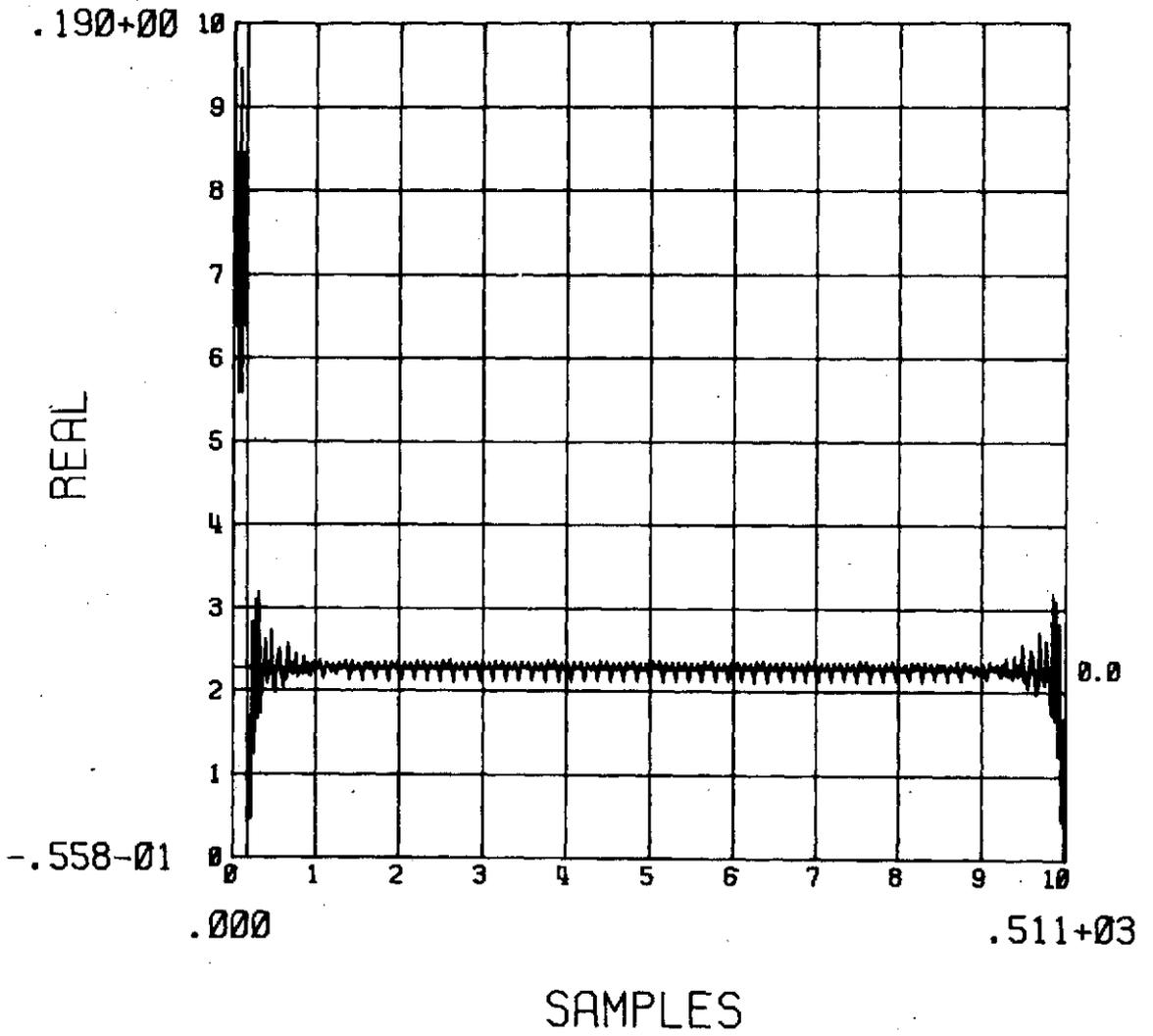


Fig. 3.6. The DDIR for LITTLE DIRSIT.

"time" sequence, so that the second half of the time sequence is the same as for negative values of n .) This result seems, prima-facie, a little odd. Note that this does not mean that "LITTLE-DIRSIT" response to an impulse before it is applied. In fact, the response of LITTLE DIRSIT to an impulse is null! This result does imply that the quasi-linear approximation to LITTLE DIRSIT will respond to an impulse before it is applied. There is, however, no a priori reason to expect a quasi-linearization of a nonlinear system to be realizable.

IV. DIRSIT

The discrete system DIRSIT is a nonlinear digital differentiator. DIRSIT has been used for some time at the University of Utah to reduce data contaminated by gaussian noise as well as random spikes. DIRSIT is a simple "hammer and tong" technique for obtaining smooth estimates of a signal and its first two derivatives. DIRSIT accepts trajectory samples and their corresponding time samples $[g(n), t(n)]$, and yields a smooth trajectory estimate, together with estimates of the velocity and acceleration at a corresponding time $[r(n), f'(n), f''(n), t(n)]$. DIRSIT utilizes a novel approach to filtering the data and providing derivative estimations. This approach is known as the derivative smoothing technique (DST). The nonlinear character of DIRSIT has made it very difficult to know what effect it has on practical signals. This situation is relieved to some extent by this research.

4.1 The Structure of DIRSIT

A complete account of the structure of DIRSIT can be found in the technical memorandum No. 1004, DA Project 516-04-007, from the White Sands Missile Range, New Mexico, November 1962. A short, slightly simplified, explanation of the structure of DIRSIT will be given here.

The conventional version of DIRSIT passed the data through an interval set of the form $[a, b), (b, c]$. The interval $[b, c]$ is referred to as the window region. The interval $[a, b)$ is referred to as

the tailoff region. Roughly, DIRSIT smooths the data by assuming a constant second derivative over the window region such that the resulting smooth data lies "in the middle" of the raw data. In particular, consider a smooth estimate of the output function $f(n)$ and its first two derivatives, $f'(n)$ and $f''(n)$, for $n = a, \dots, b - 1$. The values of $f(b)$ and $f'(b)$ are calculated based on the previous point, $b - 1$. The value of $f''(b)$ is then given an initial guess (probably $f''(b - 1)$). Then $f(n)$ is calculated based on $f''(n) = f''(b)$ for $n = b + 1, \dots, c$. If there are as many raw points above the values of $f(n)$ as there are below, then the guess for $f''(b)$ is taken to be correct. Otherwise, a new guess of $f''(b)$ is obtained, and the procedure is repeated until the "correct" value is found. The change from $f''(a)$ to $f''(b)$ is then utilized in updating the tailoff region $[a, b)$ so that there is a constant change in $f''(n)$ from $n = a$ to $n = b$. At this point, the data $f(a)$, $f'(a)$, and $f''(a)$ is output and the interval set is stepped forward to include a new data point at $n = c + 1$.

An important feature of DIRSIT is that it smooths the data via the derivative smoothing technique (DST). The DST amounts to adjusting the estimated derivative so that the resulting integrated curve "fits" the data. Thus, given $f(b)$ and $f'(b)$, $f''(b)$ is adjusted according to a prespecified model (in this case, a constant acceleration over the window), so that the resulting $f(n)$ "fits" the data. This technique provides a smooth approximation to the data because the smooth data $f(n)$ must be integrated (smoothed) twice from $f''(n)$.

Another important feature of DIRSIT is the method the DST utilizes to "fit" the data. This method is to choose $f''(b)$ such that $f(b)$ forms a "median curve" through the raw data. This method gives DIRSIT a strong immunity to noise spikes.

DIRSIT has been used in an altered form at the University of Utah. The basic structure is the same except that there is no use made of the tailoff region, i.e., the tailoff region has zero points in it. This author believes that more smoothing of the second derivative could be accomplished in a much more appropriate and less arbitrary manner than using the tailoff scheme. In this paper, unless otherwise explicitly stated, the term "DIRSIT" refers to this modified version.

The version analyzed here assumes that the window is of length 10 and that all data are sampled at equal intervals. This latter restriction is made for ease and generality of analysis. Thus, the results will be stated in such terms as "radians per sample" so that the results are applicable to all equally spaced sampling rates.

4.2 The "Frequency Response" of DIRSIT

DIRSIT is analyzed by the DDF developed in the previous chapters of this paper. The same conventions are used here that were used in the previous chapter.

The measured DDF is plotted in Fig. 4.1. The ERF is plotted in Fig. 4.2, where a maximum value of 2.0 is set on the plot so that the relevant information is clear. From this latter plot, it is clear that

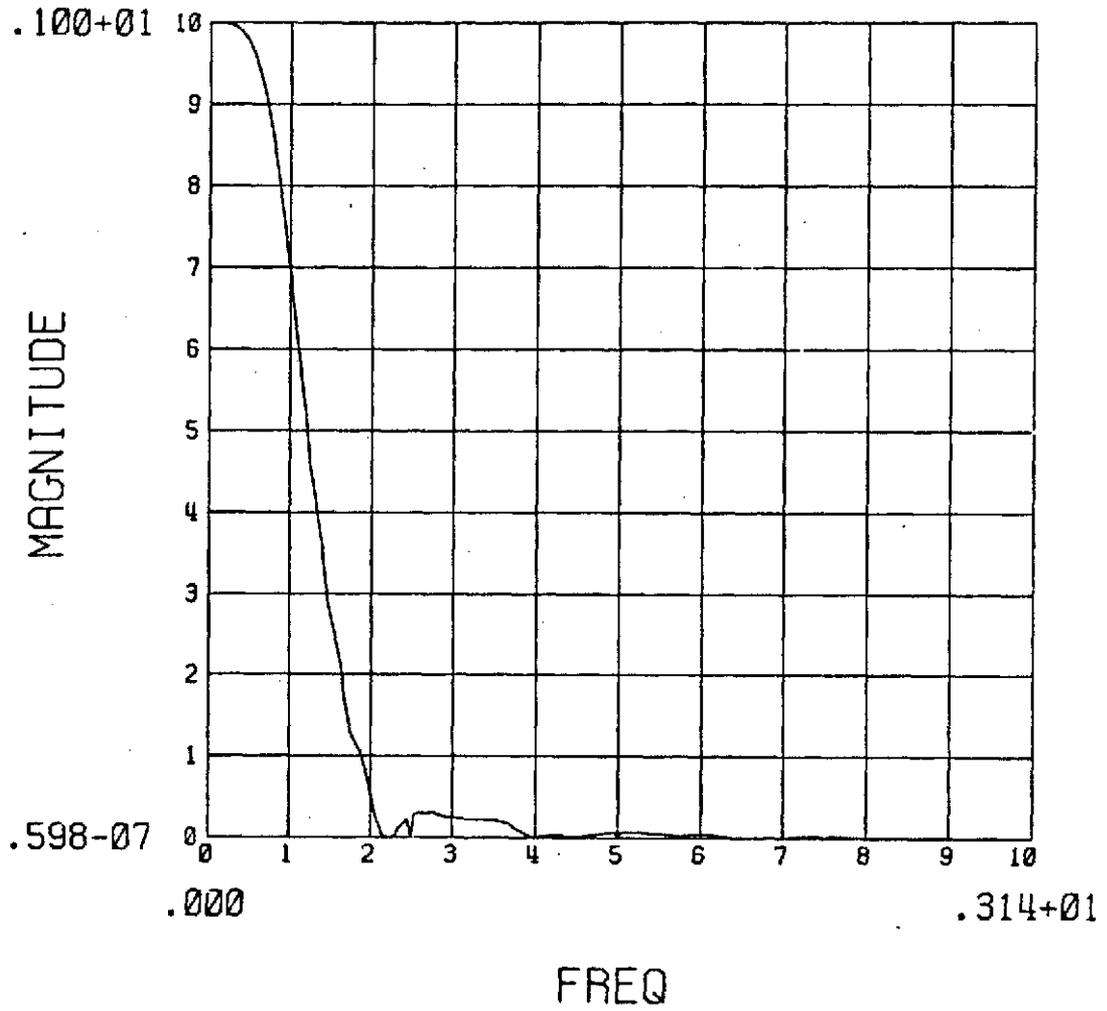


Fig. 4.1a. The measured DDF for DIRSIT. The magnitude (linear scale) is shown here.

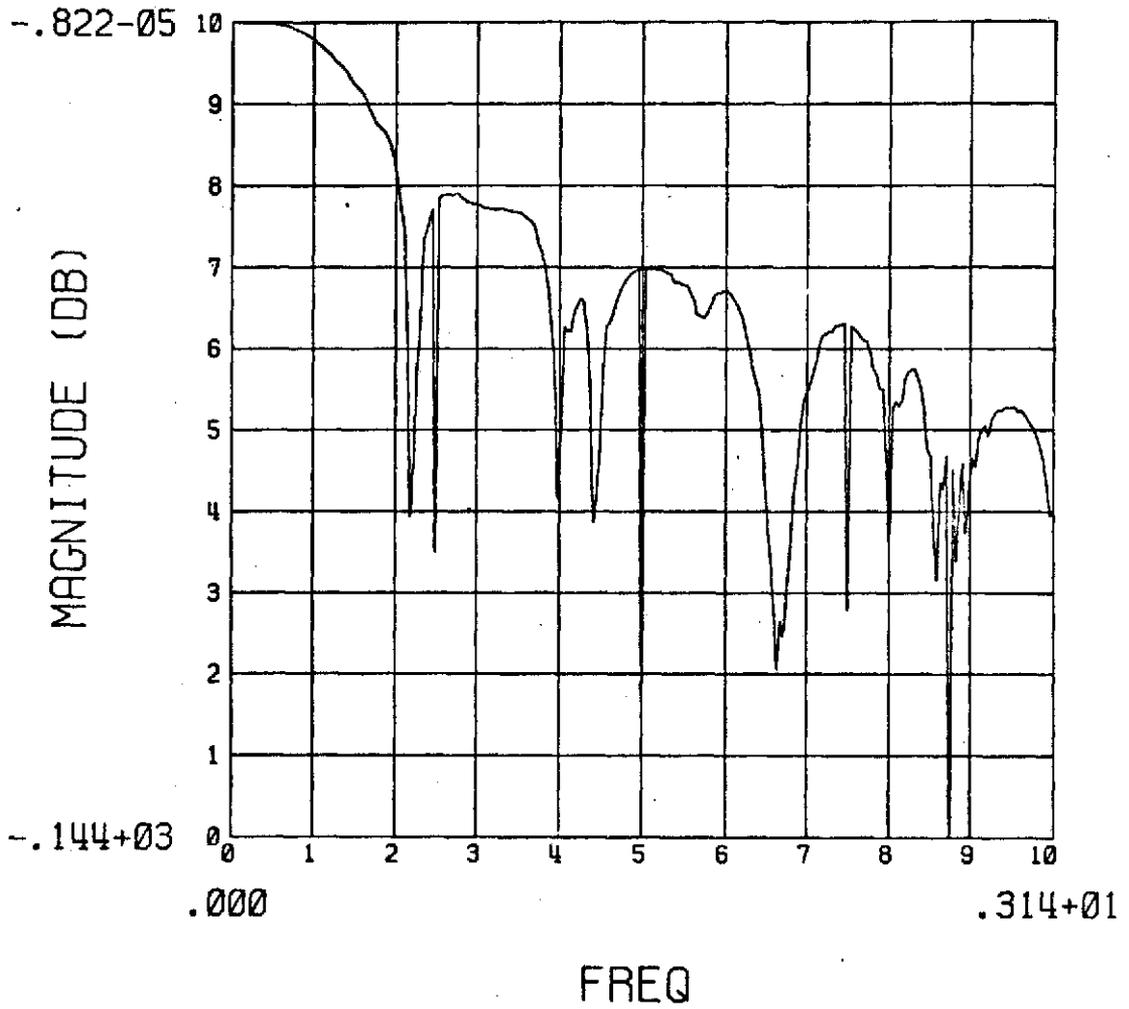


Fig. 4.1b. The measured DDF for DIRSIT. The magnitude (log scale) is shown here.

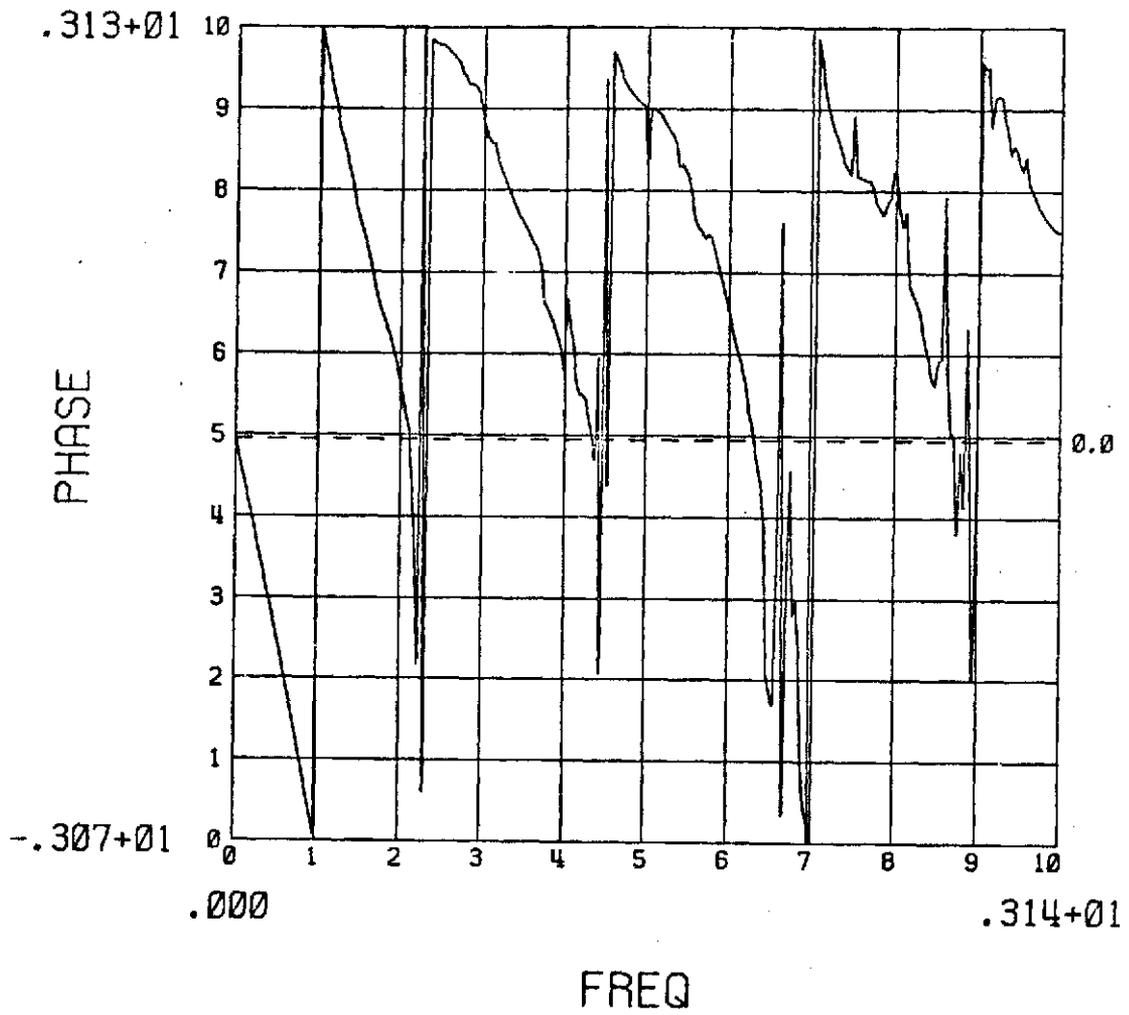


Fig. 4.1c. The measured DDF for DIRSIT. The phase is shown here.

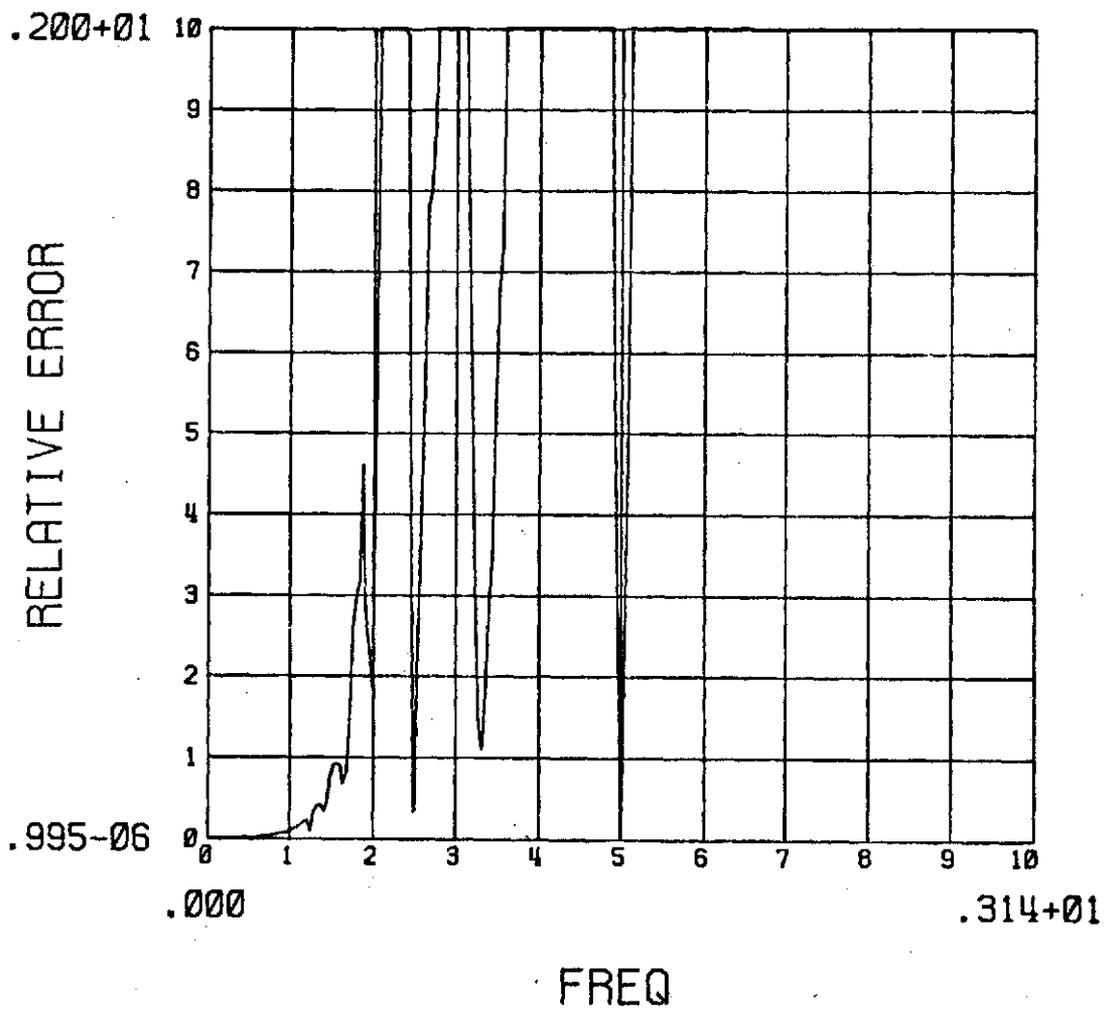


Fig. 4.2. The ERF for the DDF of DIRSIT. Note the maximum ERF is set at 2.0 to show the relevant data.

the DDF "description" is a good one only in the range from 0.0 to about 0.5 radians per sample. This means that over this interval, the linear region, DIRSIT behaves basically like a linear filter. Thus, DIRSIT is basically a low-pass filter over its linear range.

The interpretation of this data is as follows. Data with frequencies ranging from D.C. to about 0.33 radians per sample will be passed essentially unchanged. For input data, the Nyquist sampling rate, ρ radians per sample, corresponds to two samples per cycle. For data to be passed by DIRSIT, the data are specified to be below about 0.33 radians per sample, which correspond to about 20 samples per cycle. Thus, for the data to pass DIRSIT, they must be sampled about 10 times the Nyquist rate. Generally, then, DIRSIT will pass a signal with a half wavelength of about its window length or longer. This analysis is all based on Fig. 4.1.

4.3 The "Nonlinear Behavior" of DIRSIT

A FIR realized approximation to the quasi-linearization of DIRSIT is obtained. Fig. 4.3 is the real part of the IDFT of the sampled DDF obtained in Fig. 4.1. The imaginary part is negligible. This was then tailored, i.e., truncated and extended to yield the approximate DDIR (see Fig. 4.4). The Hamming, Hanning, and Fourier windows were tried to improve the results. The Fourier window worked the best. Fig. 4.5 shows the plot of the approximate DDF; this is using the data in Fig. 4.4.

DIRSIT is compared with its approximate quasi-linearization

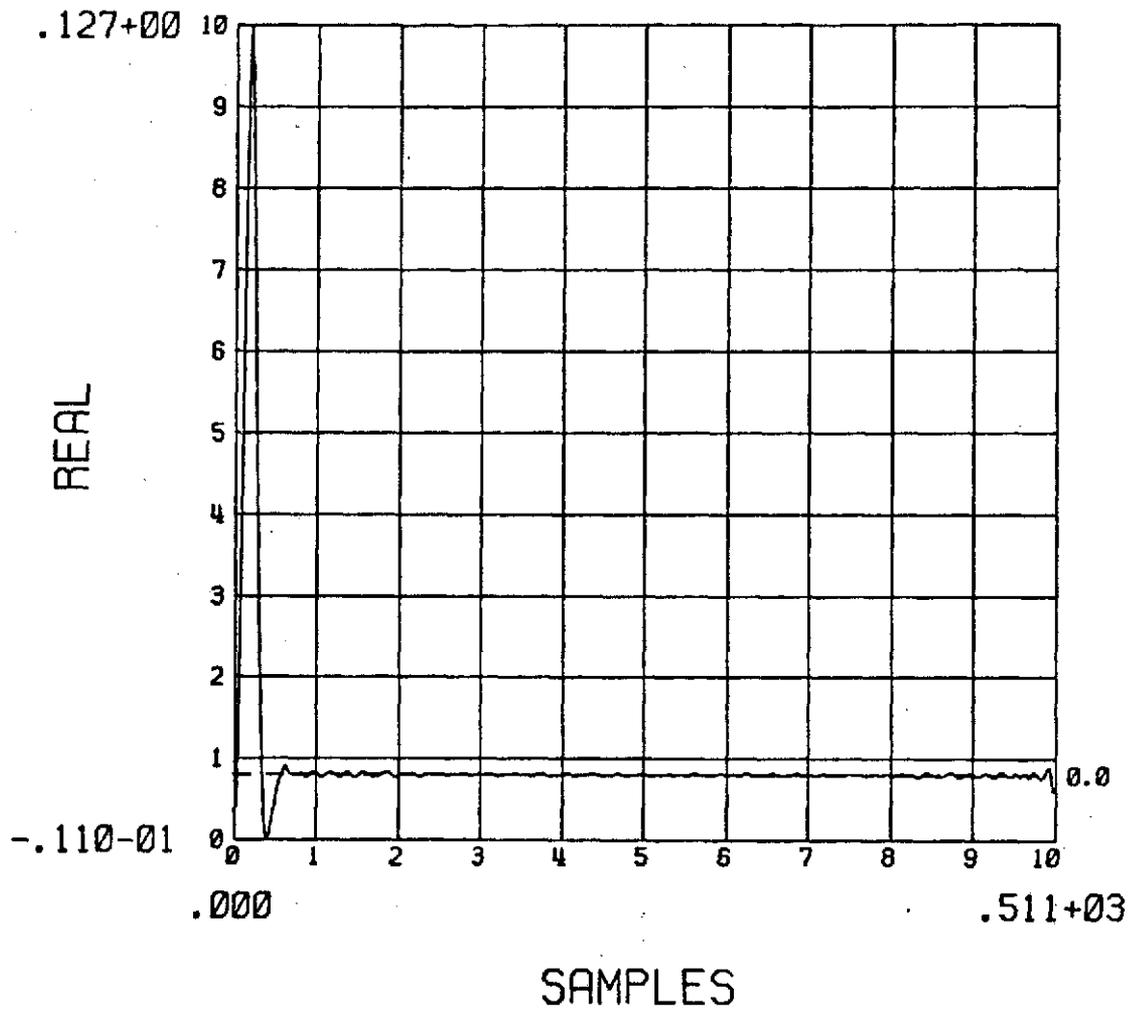


Fig. 4.3. The IDFT of the measured DDF of DIRSIT. Only the real part is shown, since the imaginary part is negligible.

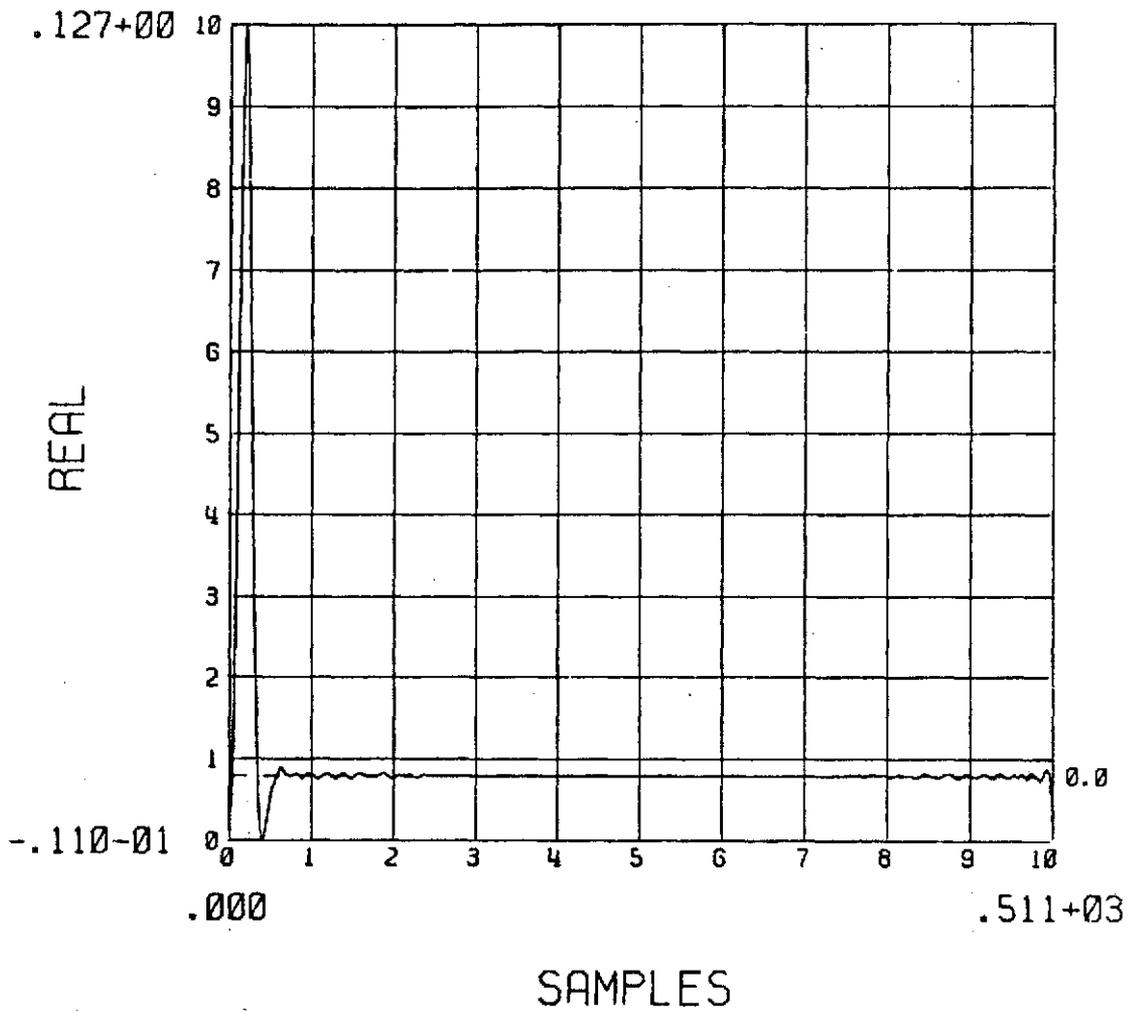


Fig. 4.4. An approximated DDIR for DIRSIT.

(see Fig. 4.4 and Fig. 4.5). This is to illustrate the "nonlinear" effects of DIRSIT. Three different inputs were fed through each system. The quality of each response to each input is based upon a qualitative judgment of the plots of the outputs. The first input is a sinusoid well within the linear region of DIRSIT. The second is the same sinusoid contaminated with gaussian noise. And the third is the noisy sinusoid contaminated again with random spikes.

The first input signal, the windowed sinusoidal, is shown in Fig. 4.6. The frequency distribution is plotted in Fig. 4.7. The windowing effect is clear from Fig. 4.7. The frequency of the sinusoidal is well within the linear region, as can be seen from Fig. 4.7 and Fig. 4.2c. The frequency of the input is 0.1 radians per sample, or about 20π samples per cycle. Thus, according to the theory thus far developed, this signal ought to be within the "pass band" of DIRSIT. (Note also that the ratio of frequency to 2π is irrational, so that this is not one of the frequencies used in measuring the DDF of DIRSIT in the first place.)

Fig. 4.8 shows the output from DIRSIT. Note the output is nearly exactly a delayed version of the input, as expected. The first and second derivatives of the input each have longer respective transient effects. The second derivative estimate (the acceleration) suffers a clipping effect.

The frequency and time plots of the quasi-linear outputs are plotted in Fig. 4.9 and Fig. 4.10, respectively. Fig. 4.10 indicates

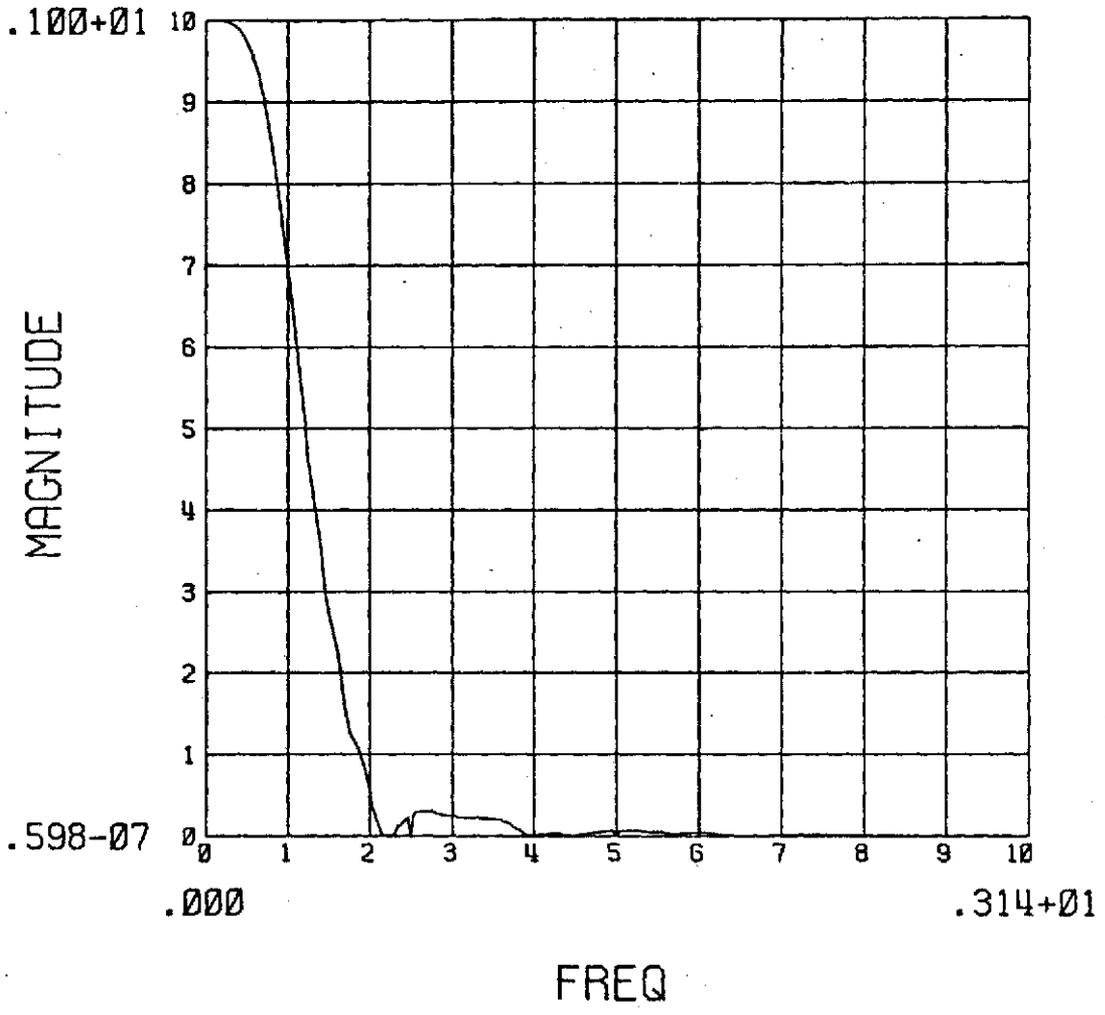


Fig. 4.5a. The approximate DDF to DIRSIT. Part (a) shows the magnitude (linear scale).

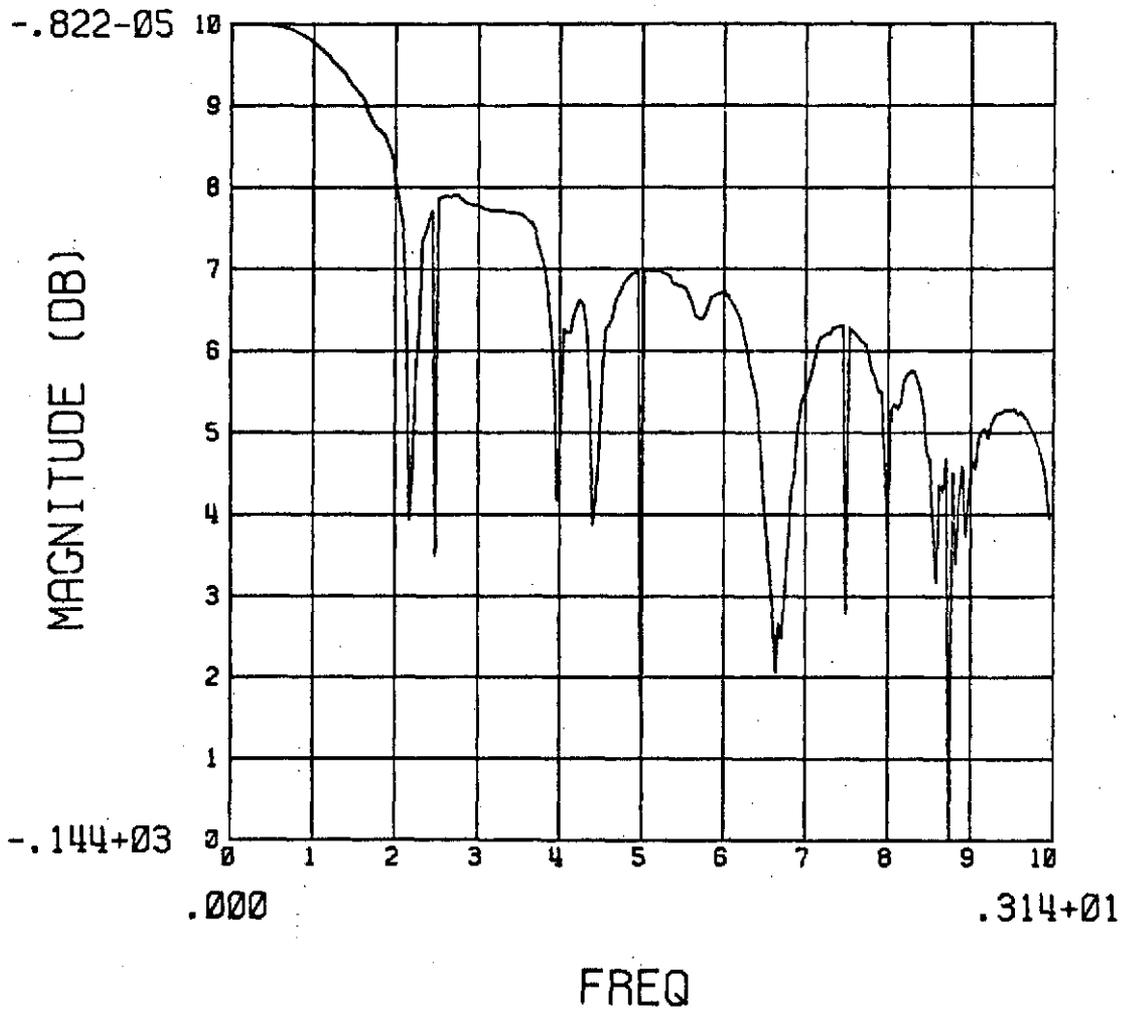


Fig. 4.5b. The approximate DDF to DIRSIT. Part (b) shows the magnitude (log scale).

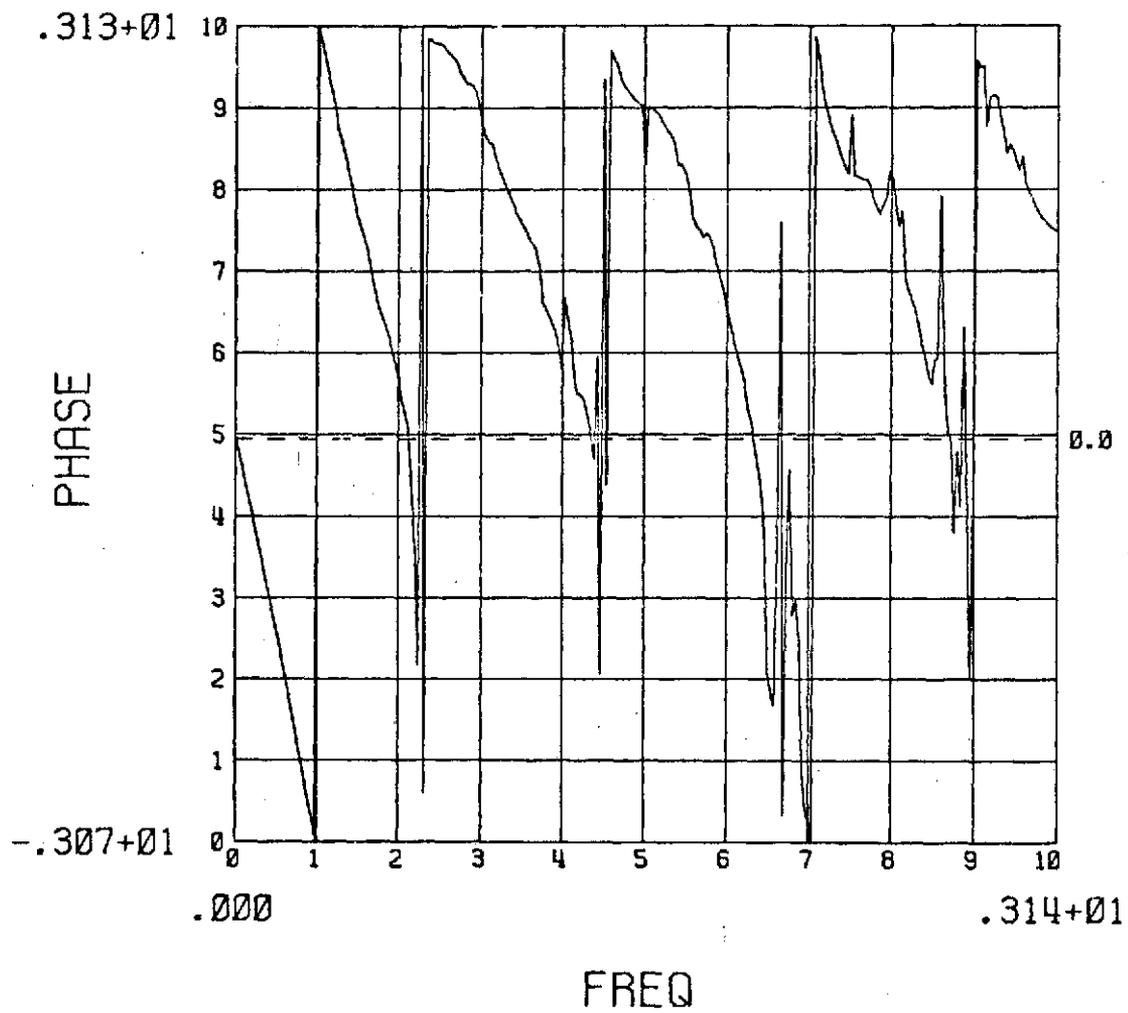


Fig. 4.5c. The approximate DDF to DIRSIT.
Part (c) shows the phase.

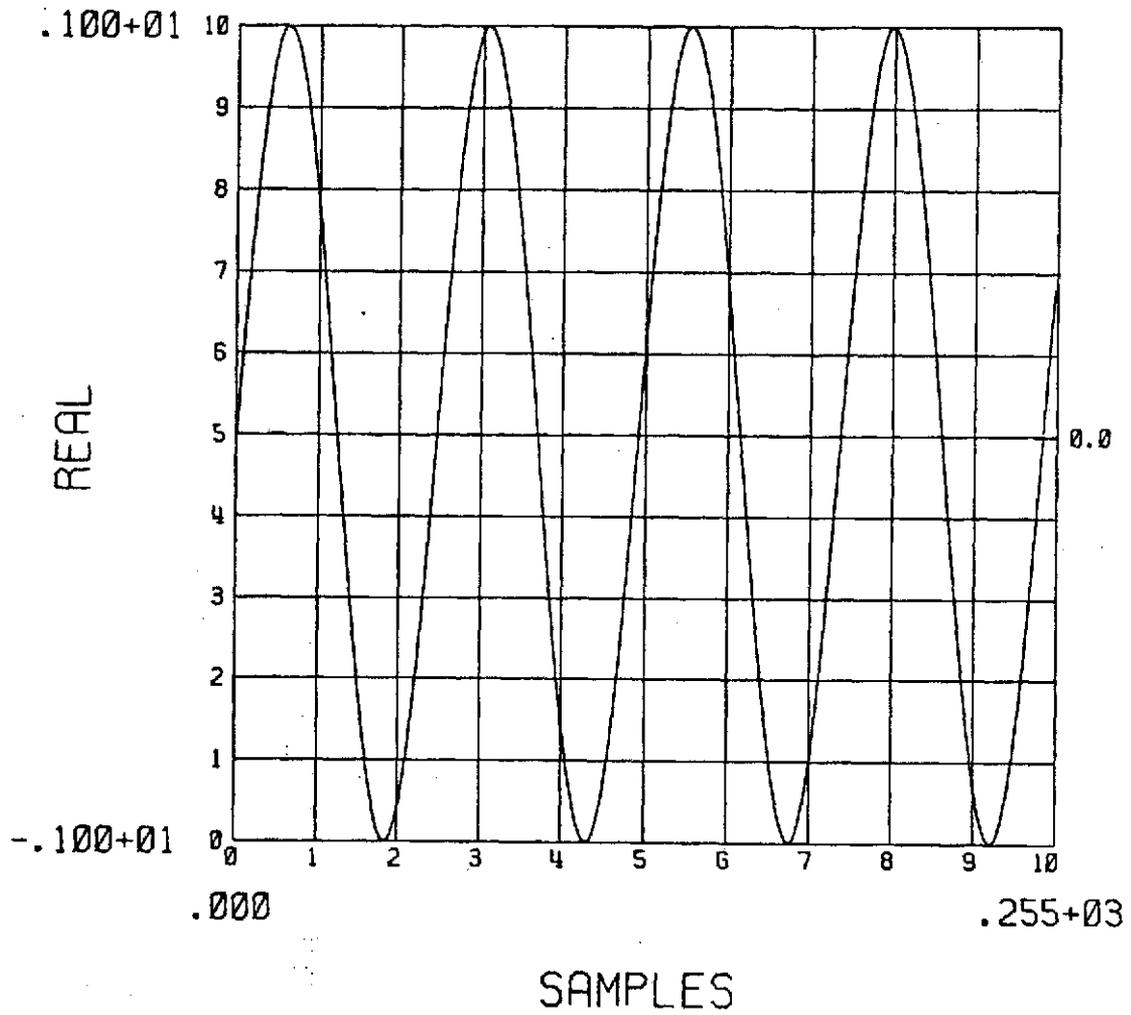


Fig. 4.6. The "first input". A windowed sinusoid with a frequency of 0.1 radians per cycle. This is the first of three inputs run through DIRSIT.

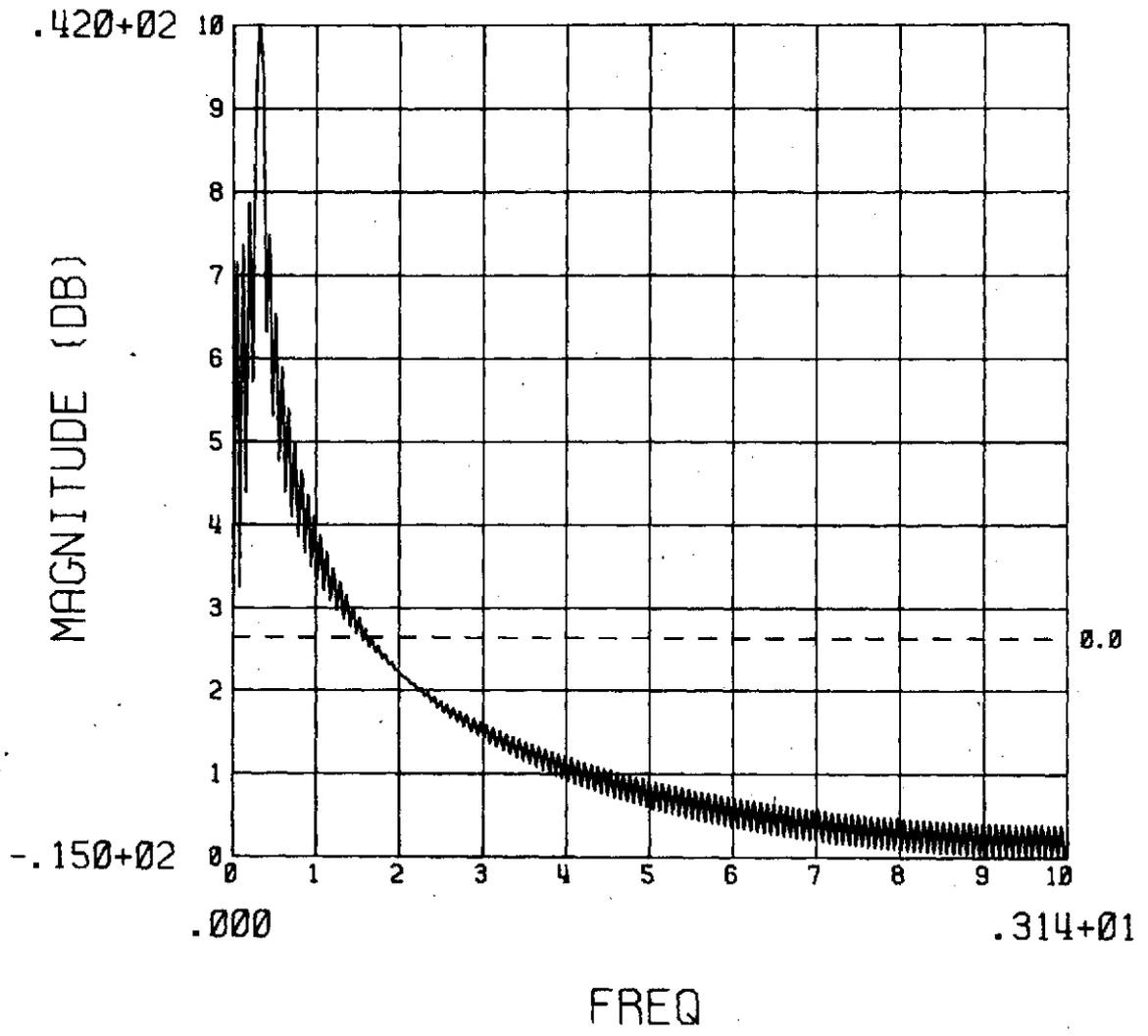


Fig. 4.7. The magnitude (log scale) of the first input shown in Fig. 4.6.

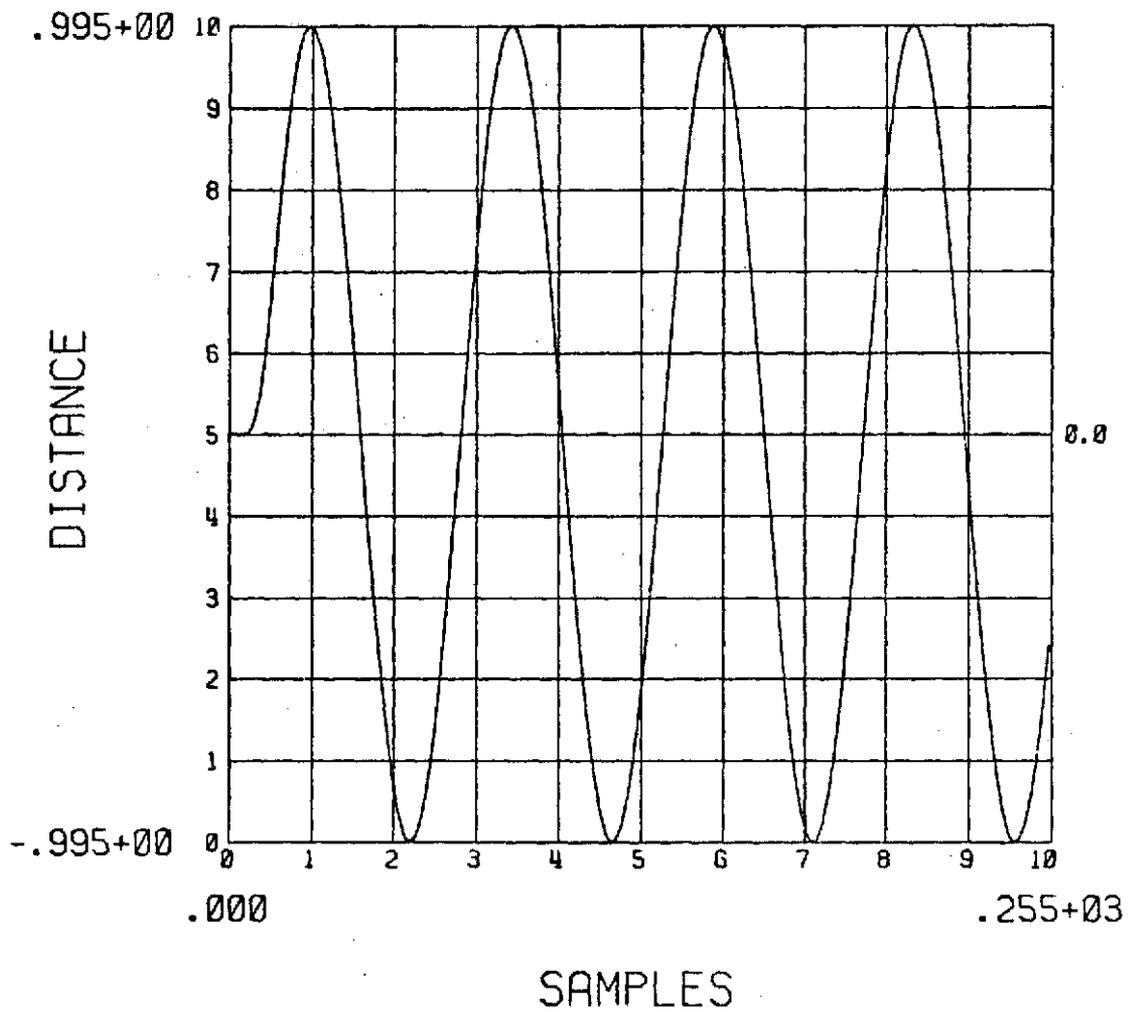


Fig. 4.8a. The response from DIRSIT to the first input. This is an estimate of the input signal (the first input).

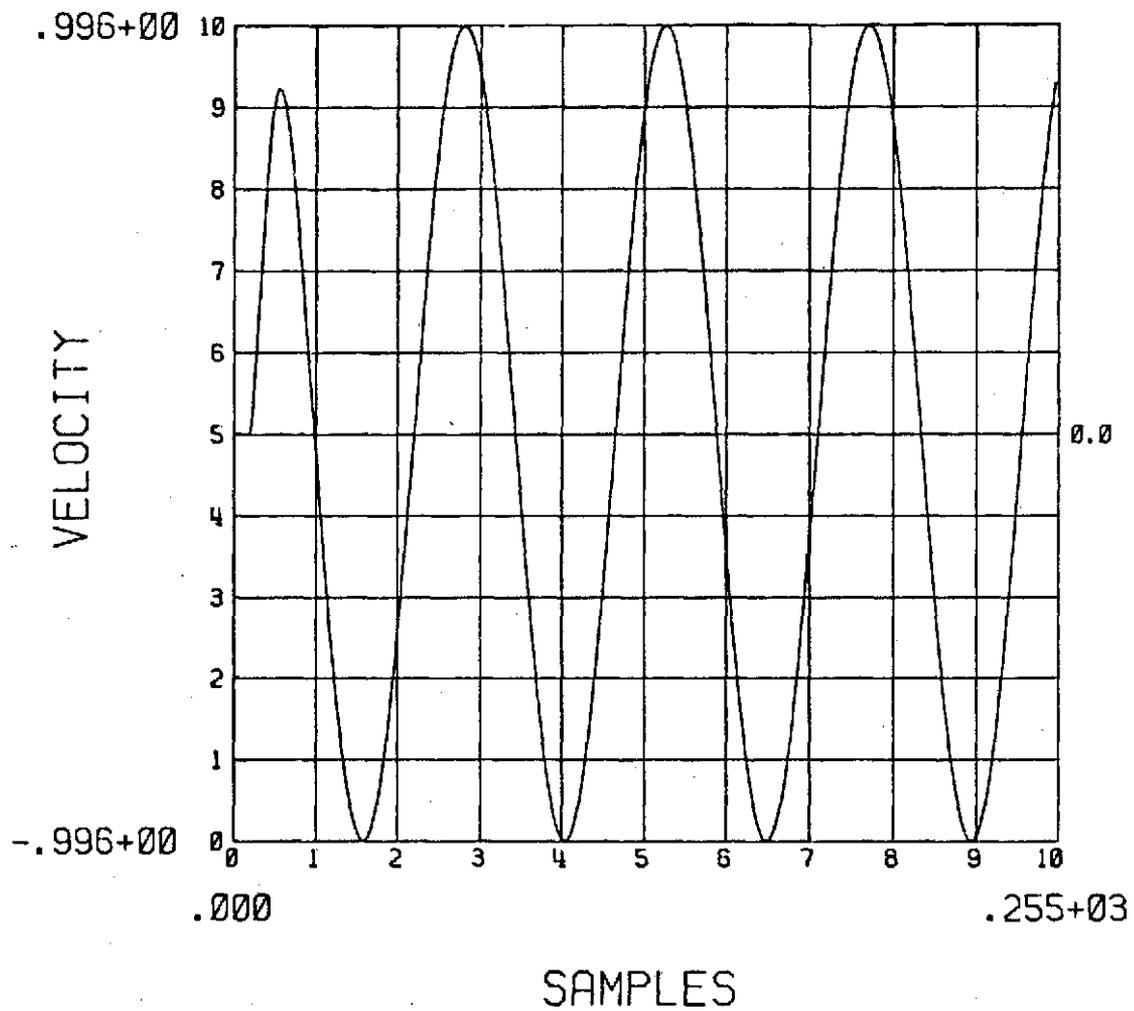


Fig. 4.8b. The response from DIRSIT to the first input. This is an estimate of the first derivative of the input signal (the first input).

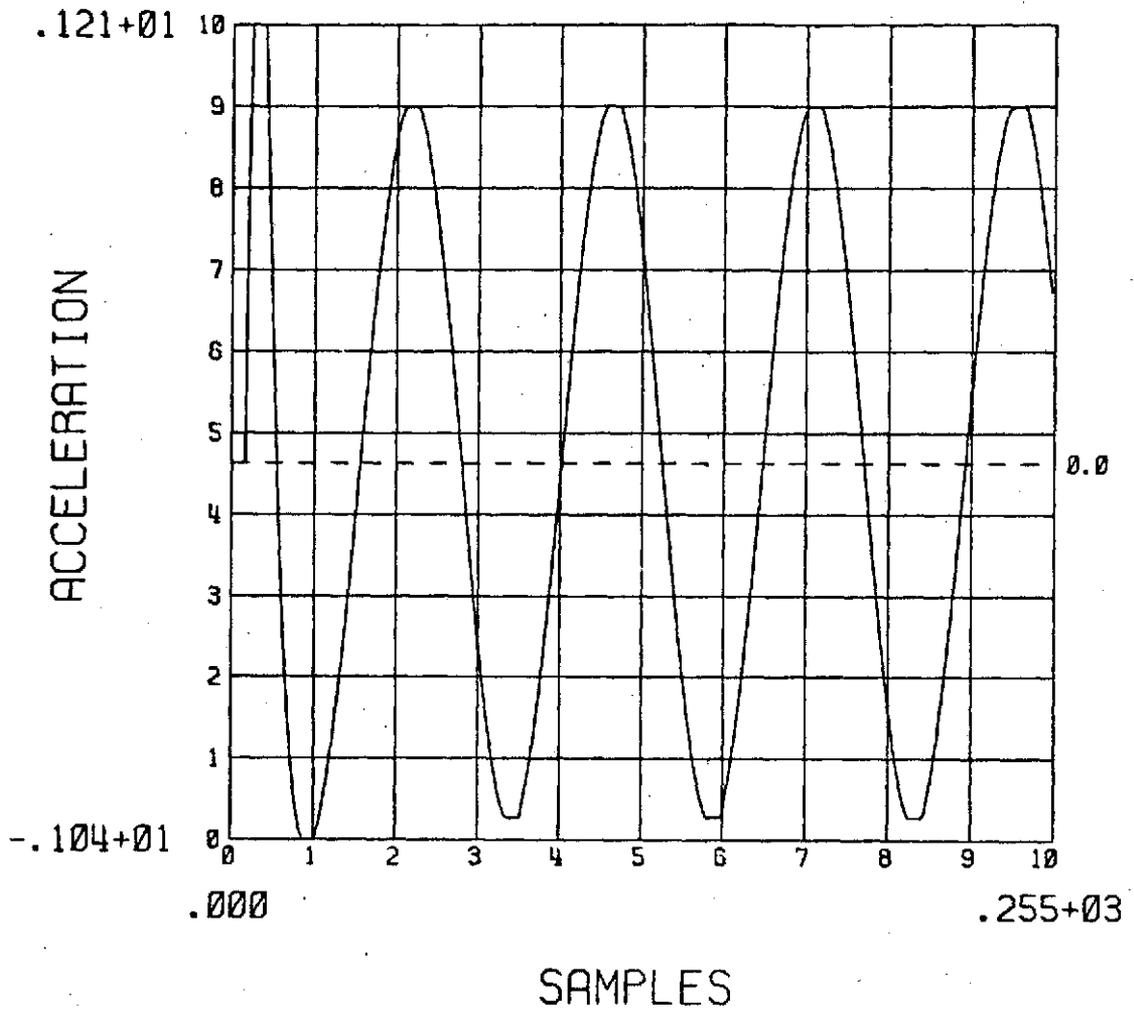


Fig. 4.8c. The response from DIRSIT to the first input. This is an estimate of the second derivative of the input signal (the first input).

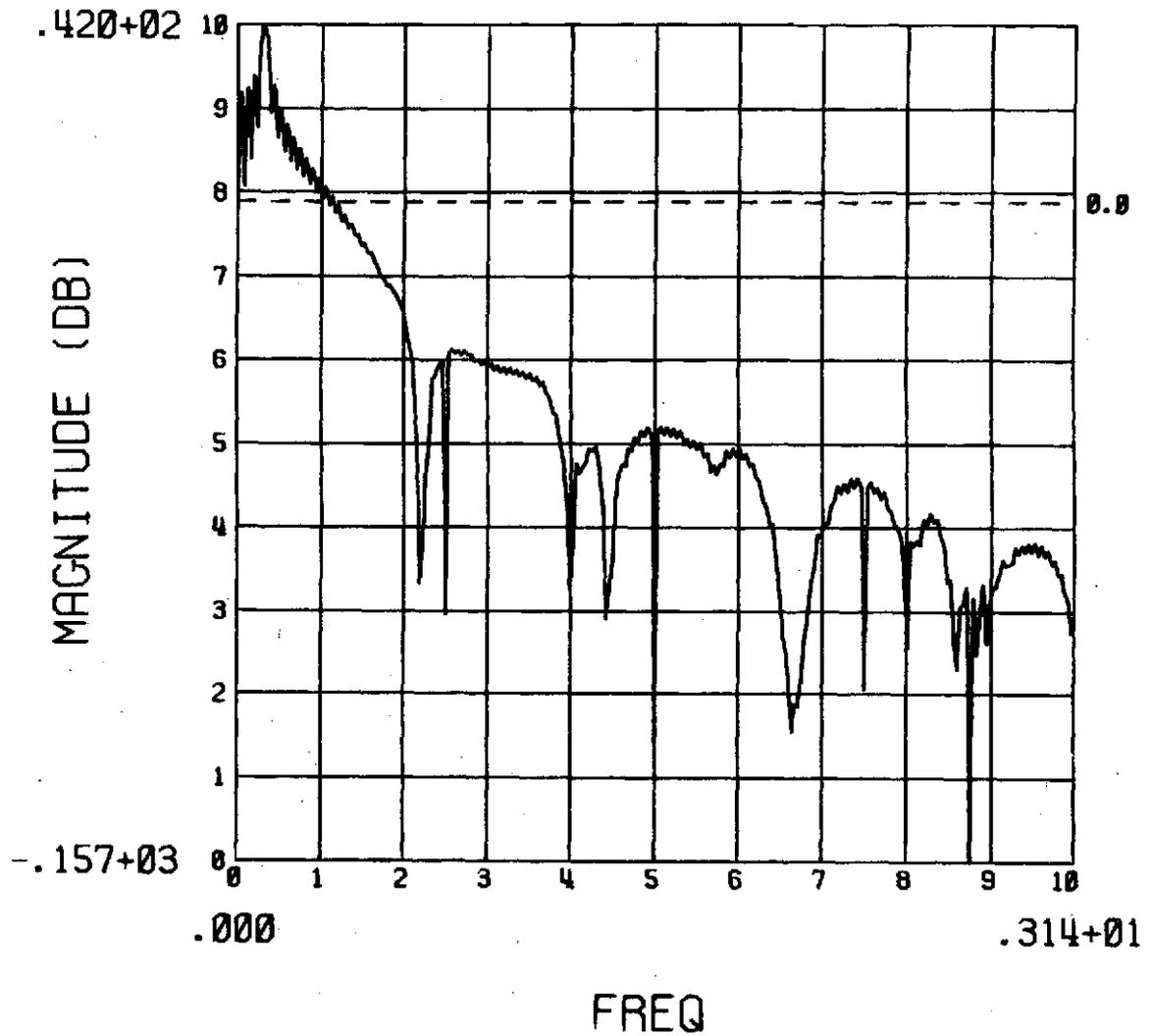


Fig. 4.9. The log magnitude of the response from the quasi-linear approximation to the first input (the frequency description).

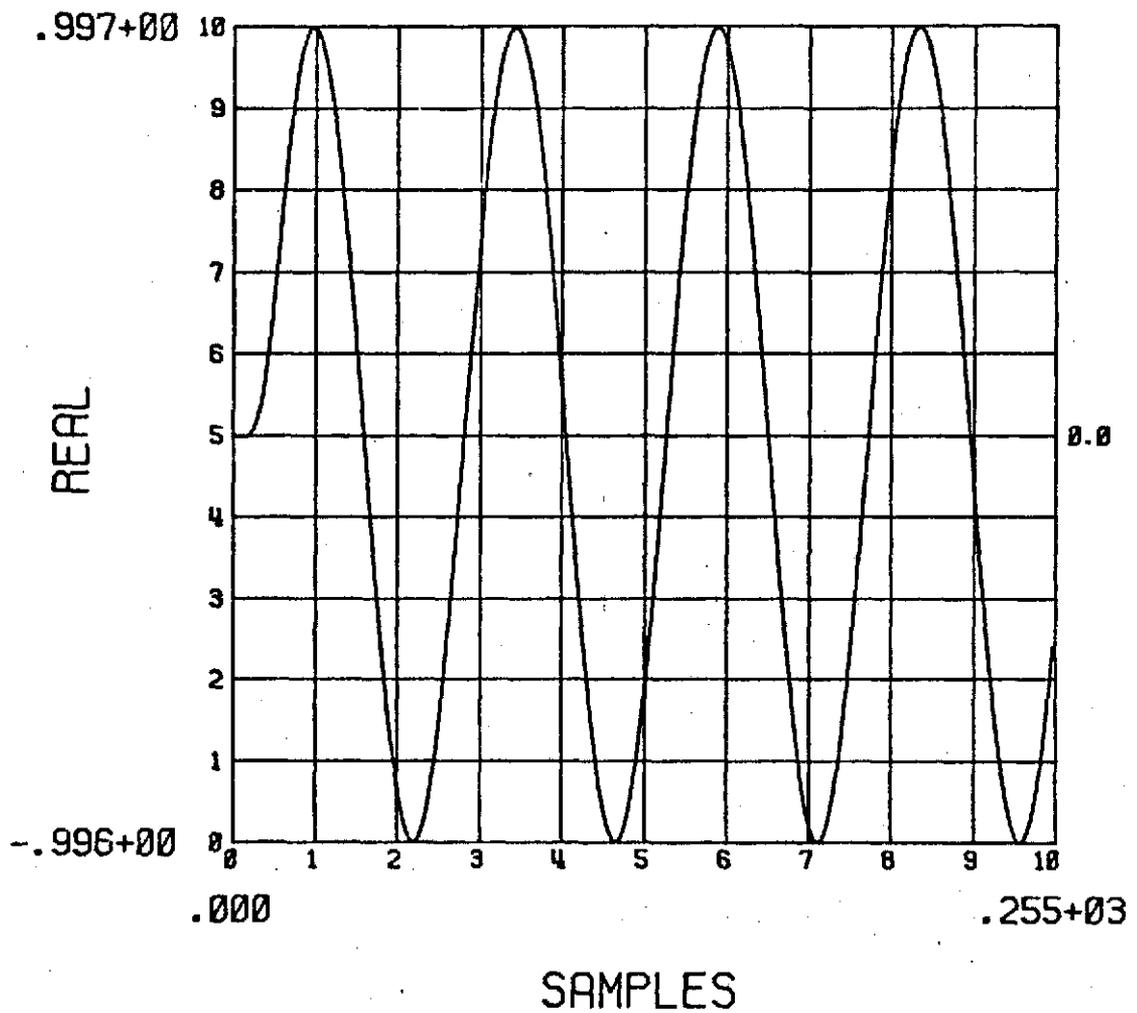


Fig. 4.10. The response from the quasi-linear approximation to the first input (the time description). Only the real part is shown, since the imaginary part is negligible.

that the quasi-linearization passes the sampled sinusoid with a slight delay. Fig. 4.11 shows the difference between the output of DIRSIT and the quasi-linearization output. After transients, the magnitude of this difference is less than 0.0015, or the outputs differ less than 0.15 percent.

The second input signal consists of the previous input signal, the sampled sinusoid, contaminated by added gaussian noise. The added noise has a mean of 0.0 and a standard deviation 0.1. This input signal is plotted in Fig. 4.12. The frequency of this signal is plotted in Fig. 4.13.

The response from DIRSIT to this second input is shown in Fig. 4.14. The smooth estimate, $f(n)$, of the noisy input seems to be affected little. However, the first derivative (velocity) shows more effect of the noise. Finally, the second derivative is quite irregular and choppy.

The frequency description and time description of the response to the quasi-linearization are shown in Fig. 4.15 and Fig. 4.16, respectively. Clearly, most of the noise is attenuated with a relatively good estimate of the sampled sinusoidal as an output.

Fig. 4.17 shows the difference between the output of DIRSIT and its quasi-linearization. This difference, after transients, is bound by 0.05. This shows that, as far as the smooth estimate of the sinusoidal goes, DIRSIT still behaves very much like a linear, low-pass filter. This result is important, since the frequency of the input

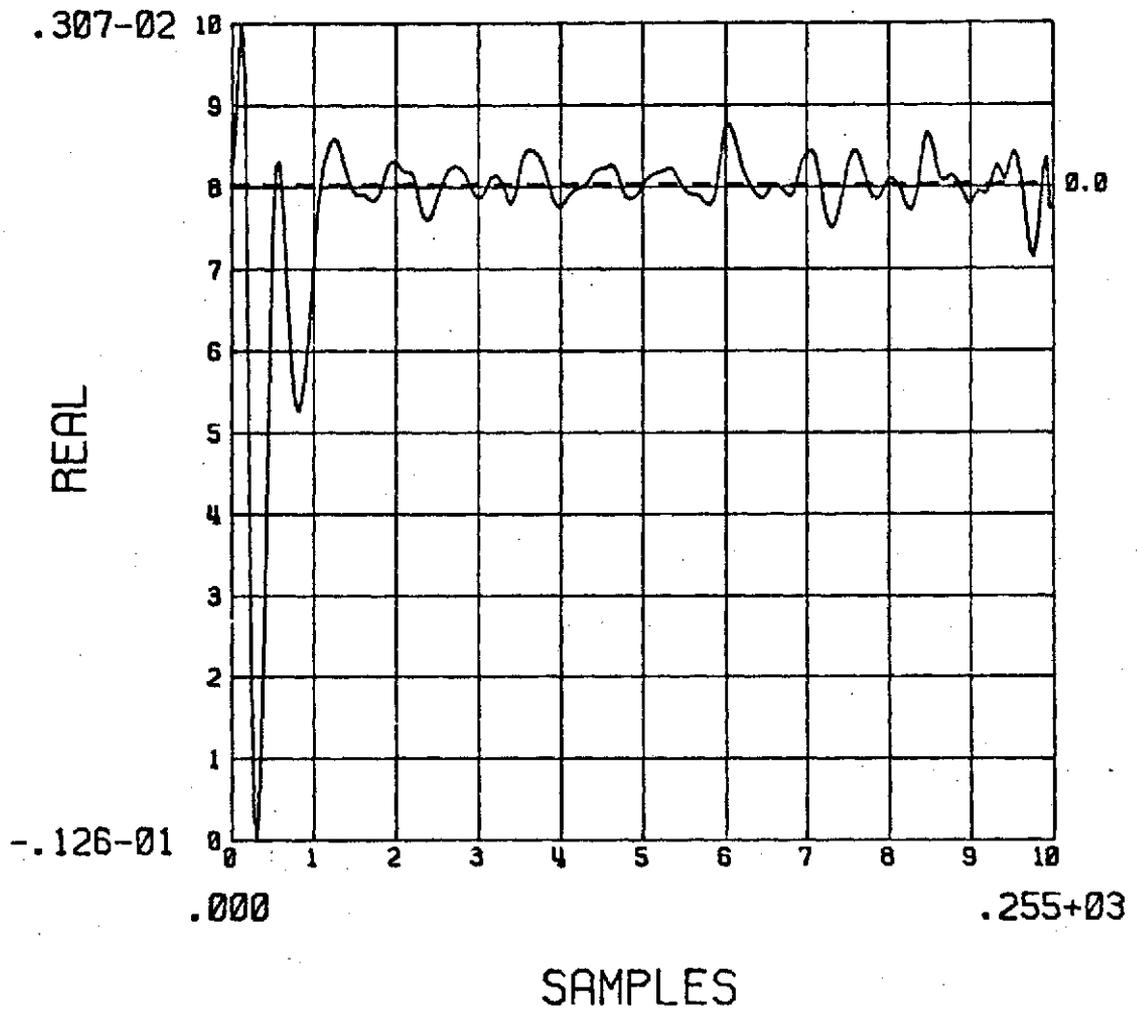


Fig. 4.11. The difference between the response from DIRSIT and the response from the quasi-linear approximation to the first input.

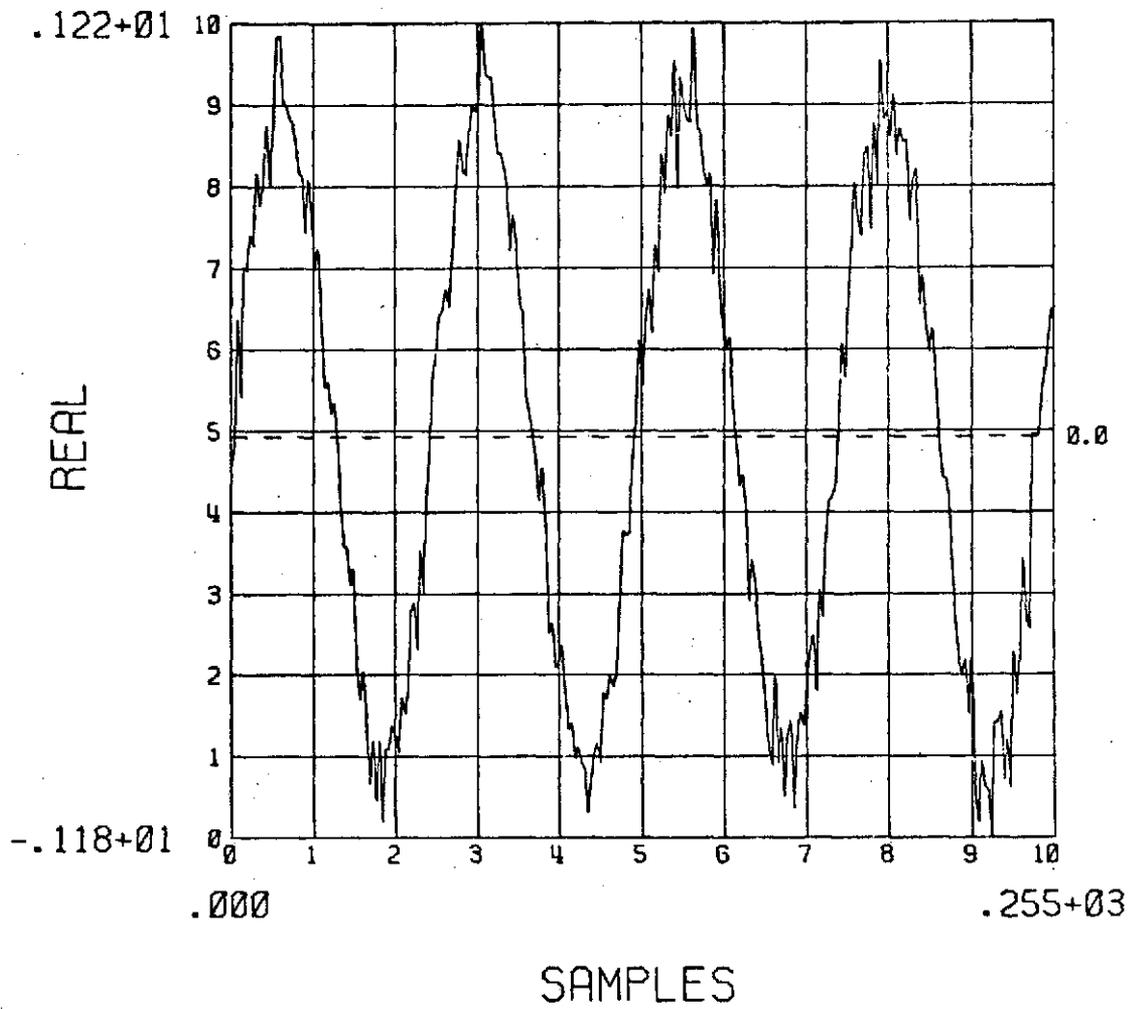


Fig. 4.12. The "second input". This is a version of the signal (the first input) contaminated with gaussian noise.

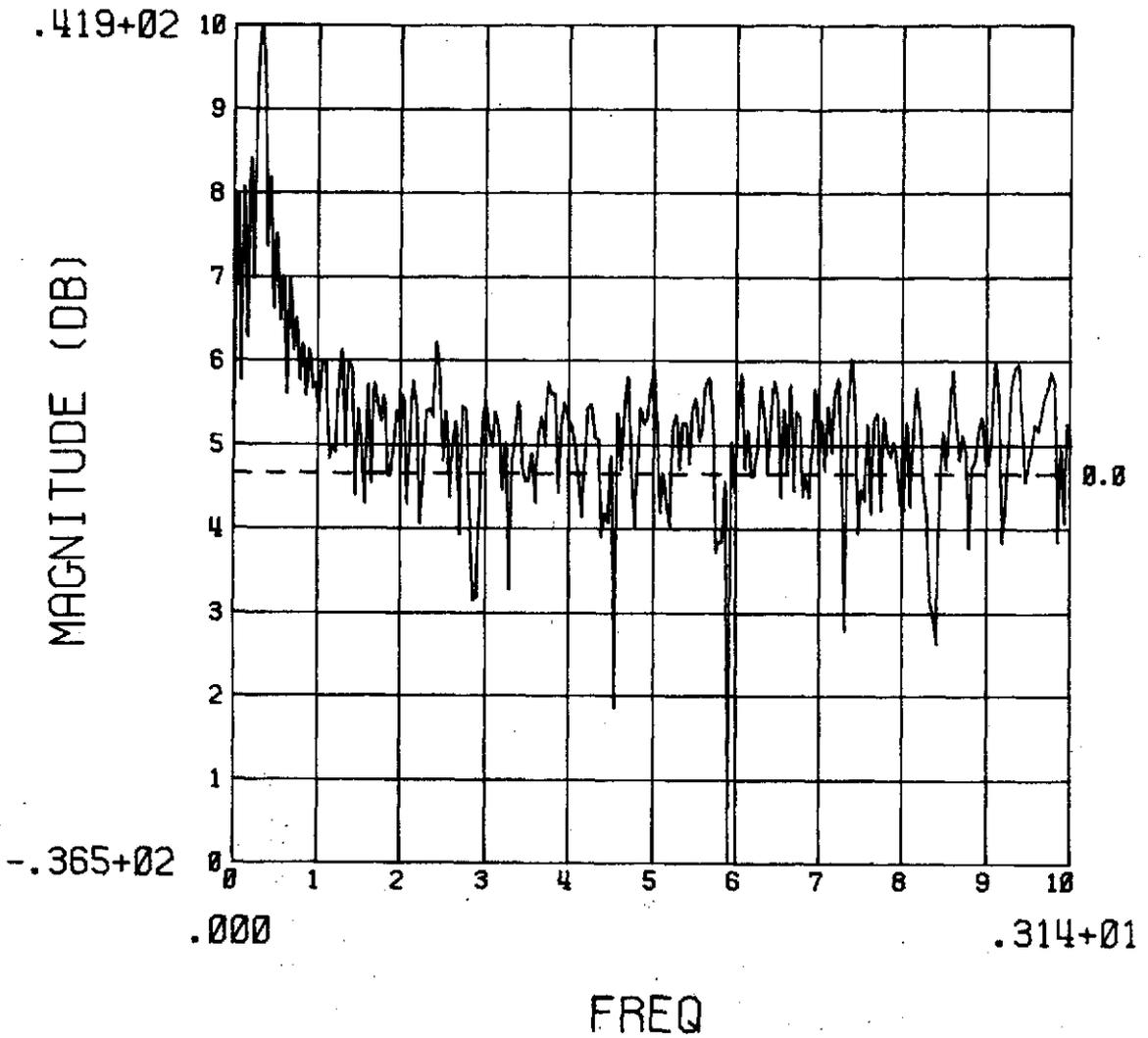


Fig. 4.13. The magnitude (log scale) of the second input shown in Fig. 4.12.

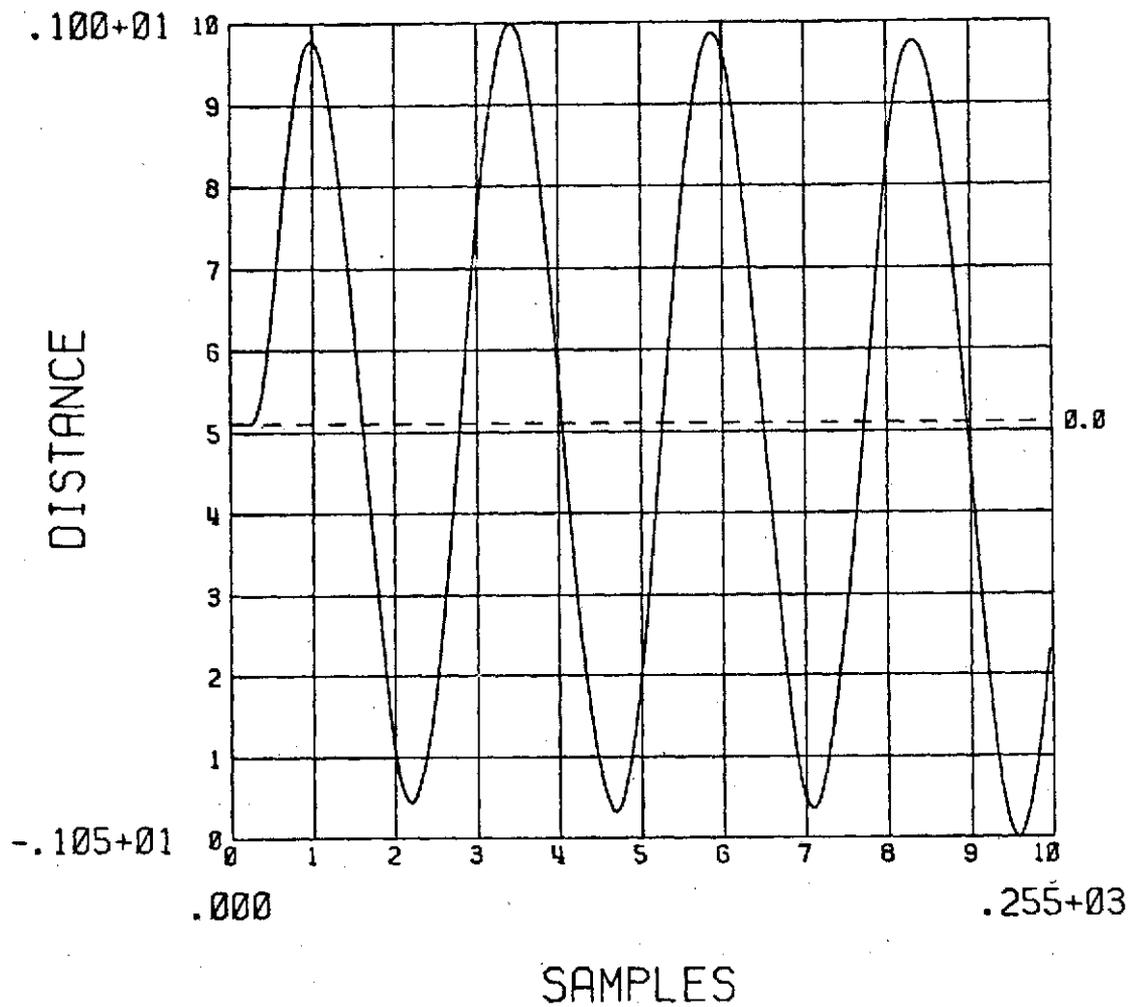


Fig. 4.14a. The response from DIRSIT to the second input. This is an estimate of the input signal (the first input).

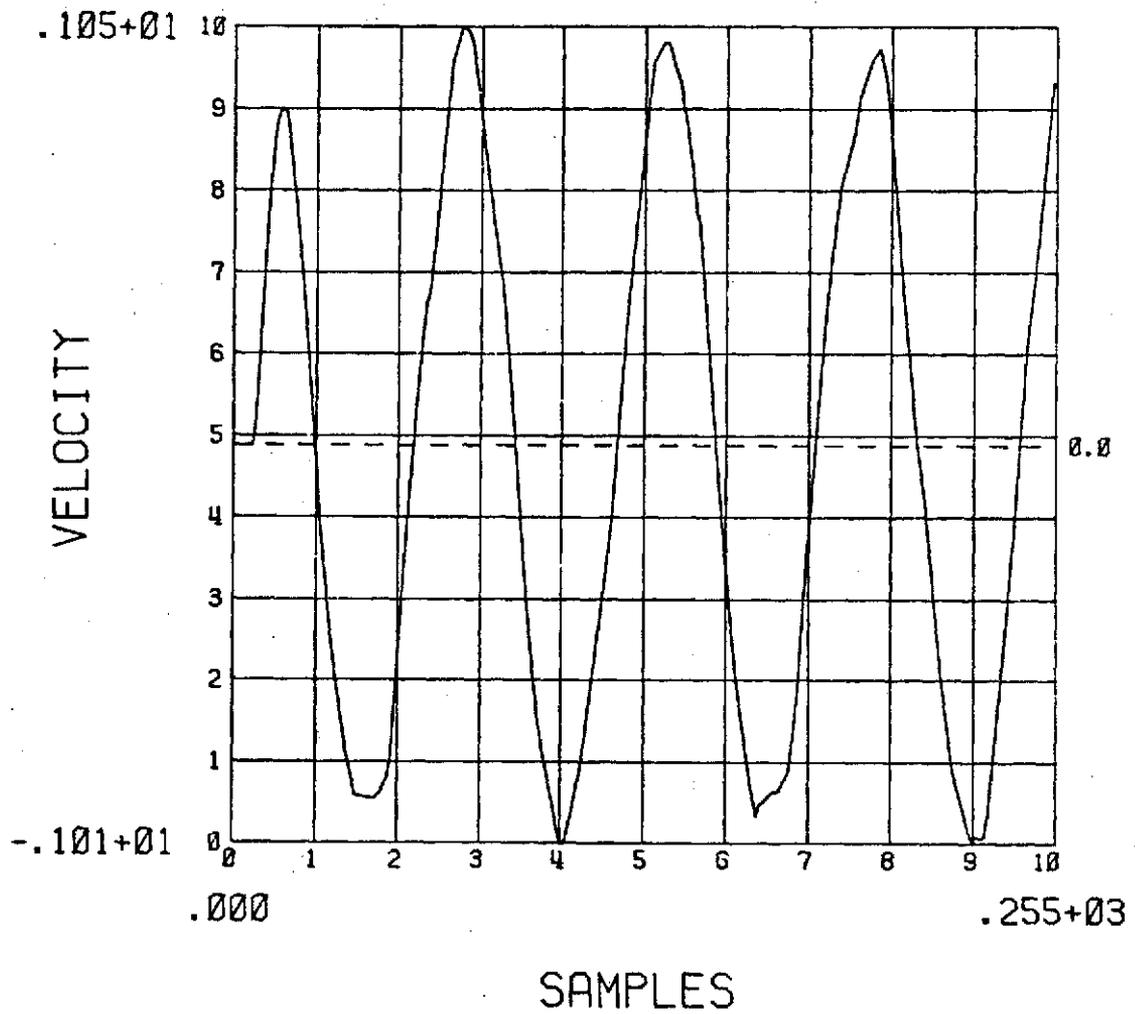


Fig. 4.14b. The response from DIRSIT to the second input. This is an estimate of the first derivative of the input signal (the first input).

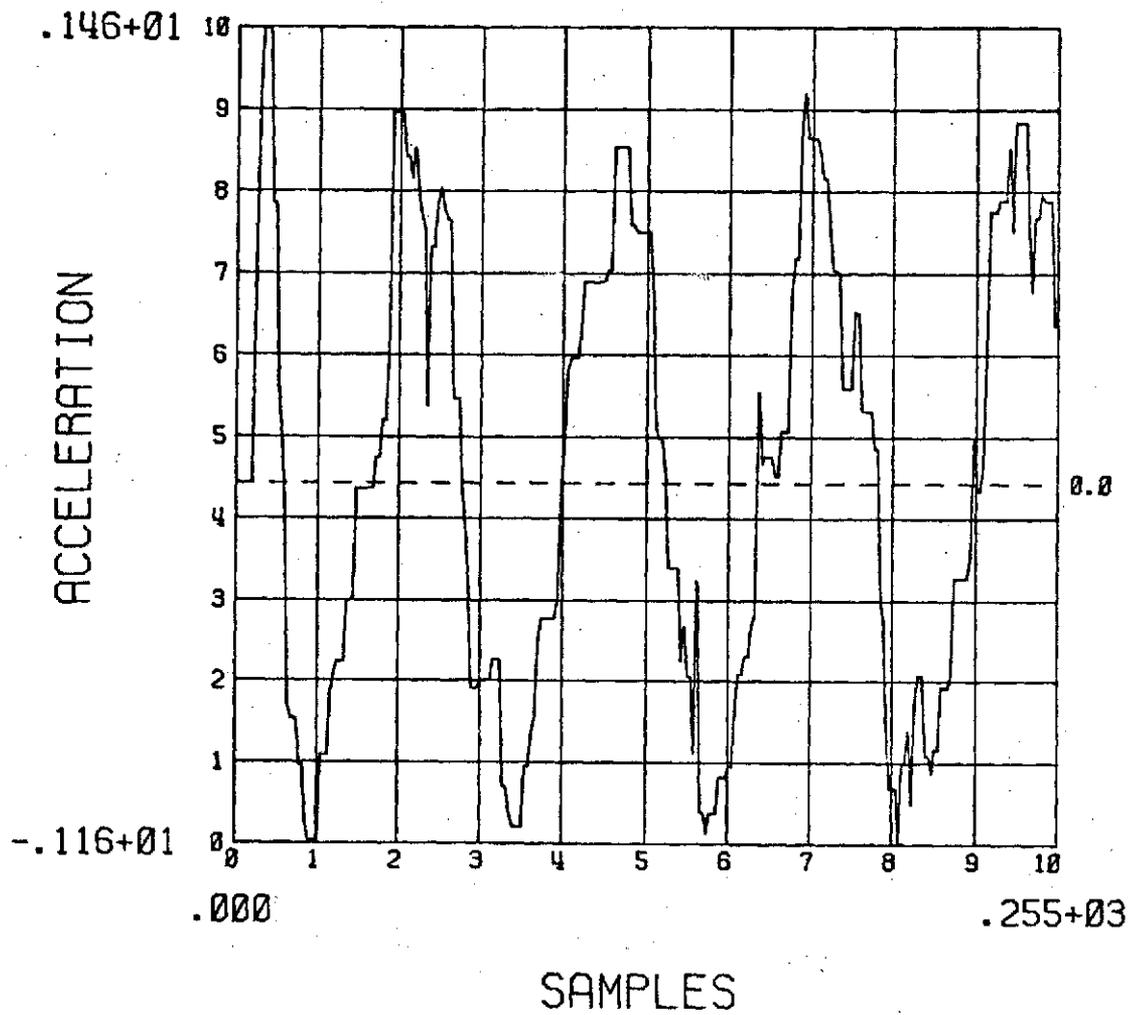


Fig. 4.14c. The response from DIRSIT to the second input. This is an estimate of the second derivative of the input signal (the first input).

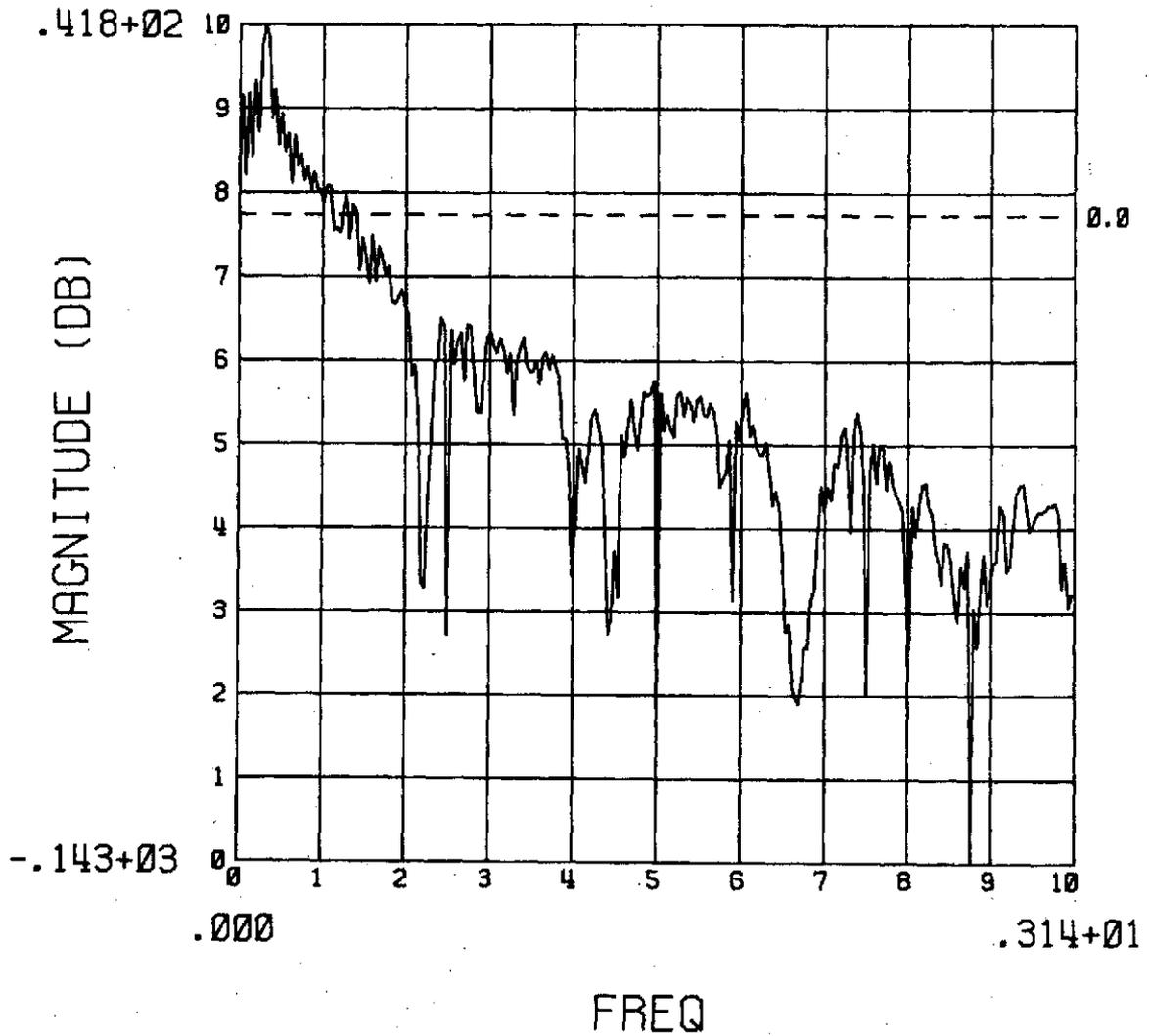


Fig. 4.15. The log magnitude of the response from the quasi-linear approximation to the second input (the frequency description).

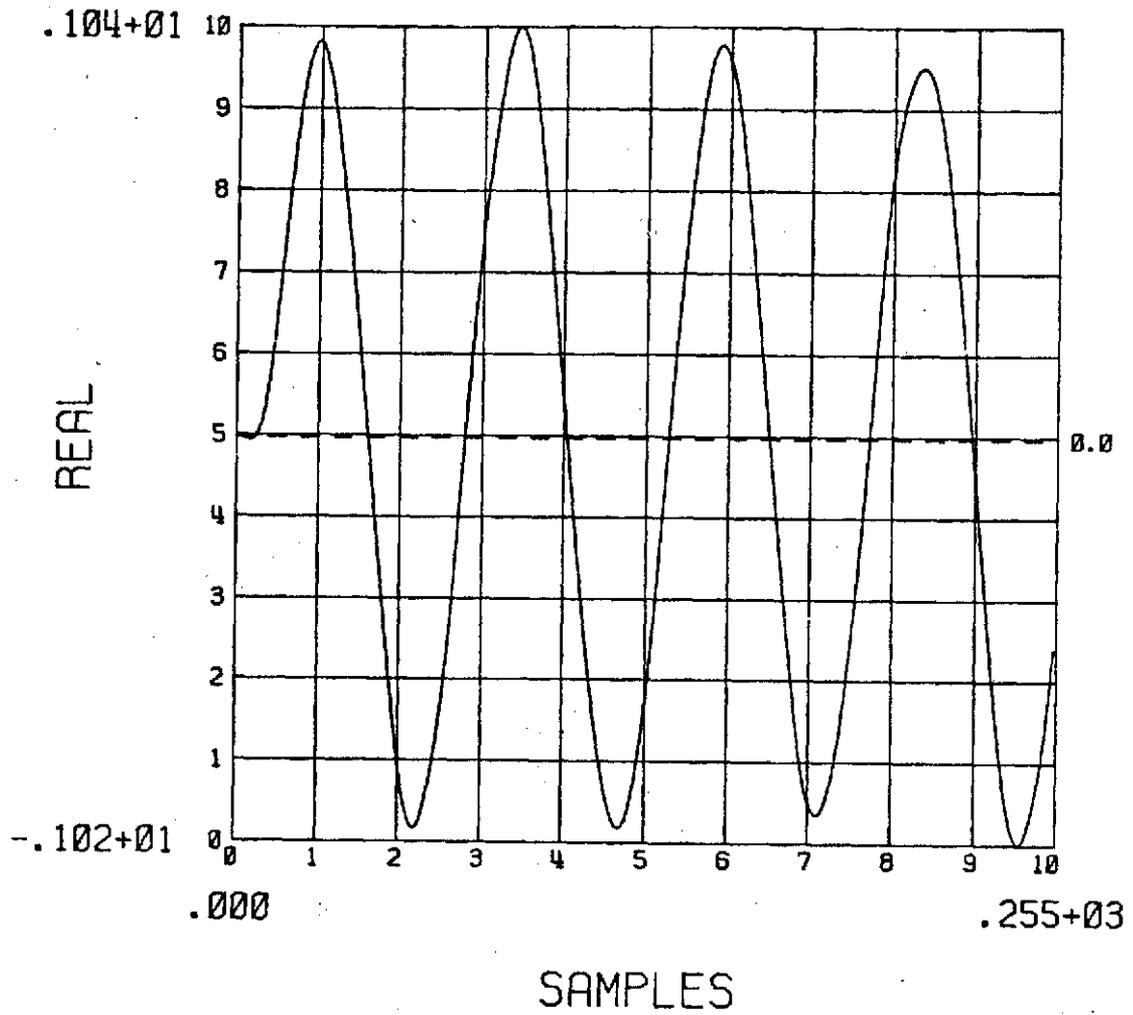


Fig. 4.16. The response from the quasi-linear approximation to the second input (the time description). Only the real part is shown, since the imaginary part is negligible.

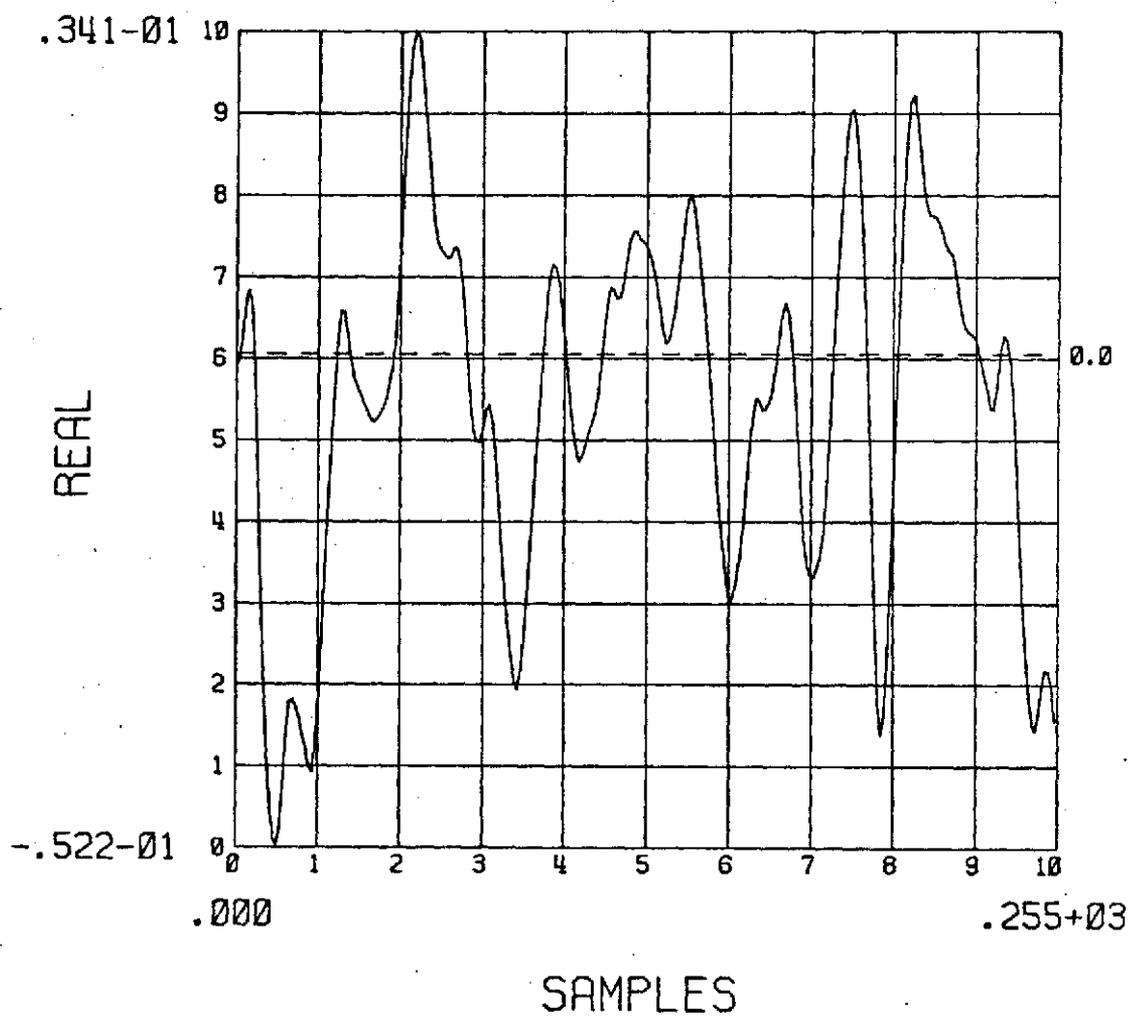


Fig. 4.17. The difference between the response from DIRSIT and the response from the quasi-linear approximation to the second input.

function is not negligible over the nonlinear range of DIRSIT (See Fig. 4.13).

The third input consists of the noisy sampled sinusoidal input contaminated with random spikes. This is plotted in Fig. 4.18. The spikes are an added constant of 5.0, added in random intervals. Thus, a bias is shown in the frequency description of this input in Fig. 4.19.

The response of DIRSIT to this input is plotted in Fig. 4.20. The results of this reduction compared to the results of the previous input is very interesting. There are two notable points. First, the smooth estimate of the function and its first derivative are about the same, and second, the second derivatives differ. The first observation implies that DIRSIT, as a digital filter, is practically immune to spikes. This immunity carries into the first derivative estimates. The second observation seems reasonable in light of the fact that the second derivative at b , $f''(b)$, is directly dependent on the latest input value at c , while $f(b)$ and $f'(b)$ are both only indirectly dependent on the point at c . Thus, a sort of built in "smoothing effect" is in practice. This is another result of the DST. Examination of Fig. 4.19 and Fig. 4.13 shows the noise is much more prevalent, so that the quasi-linearization is expected to fair much worse. This is indeed the case as can be seen from Fig. 4.12 and Fig. 4.22.

The quasi-linearization response to this third input shows a large discrepancy from both the original sinusoid and for the smooth

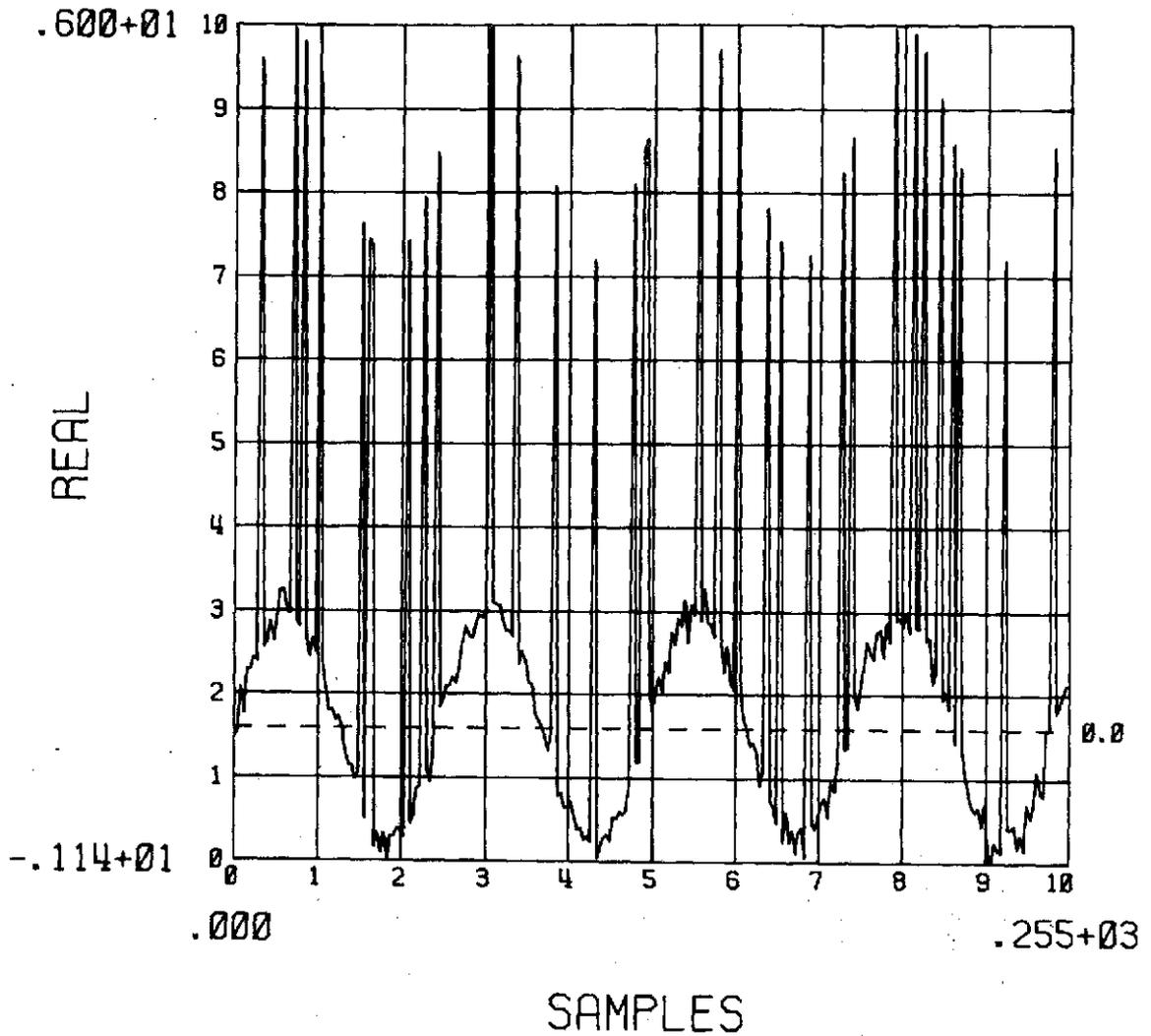


Fig. 4.18. The "third input". A version of the noisy input signal (the second input) contaminated again with spikes.

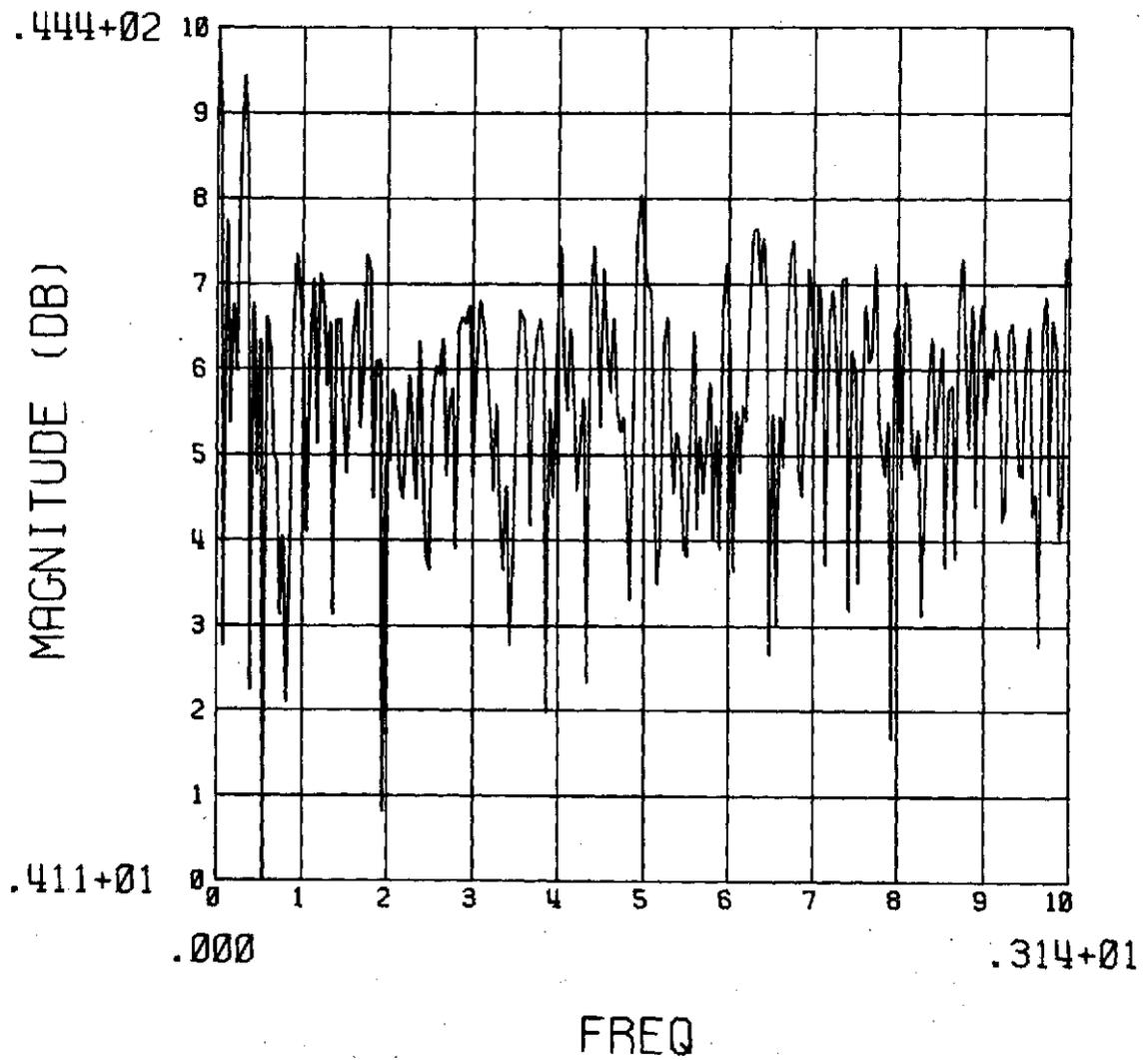


Fig. 4.19. The magnitude (log scale) of the third input shown in Fig. 4.18.

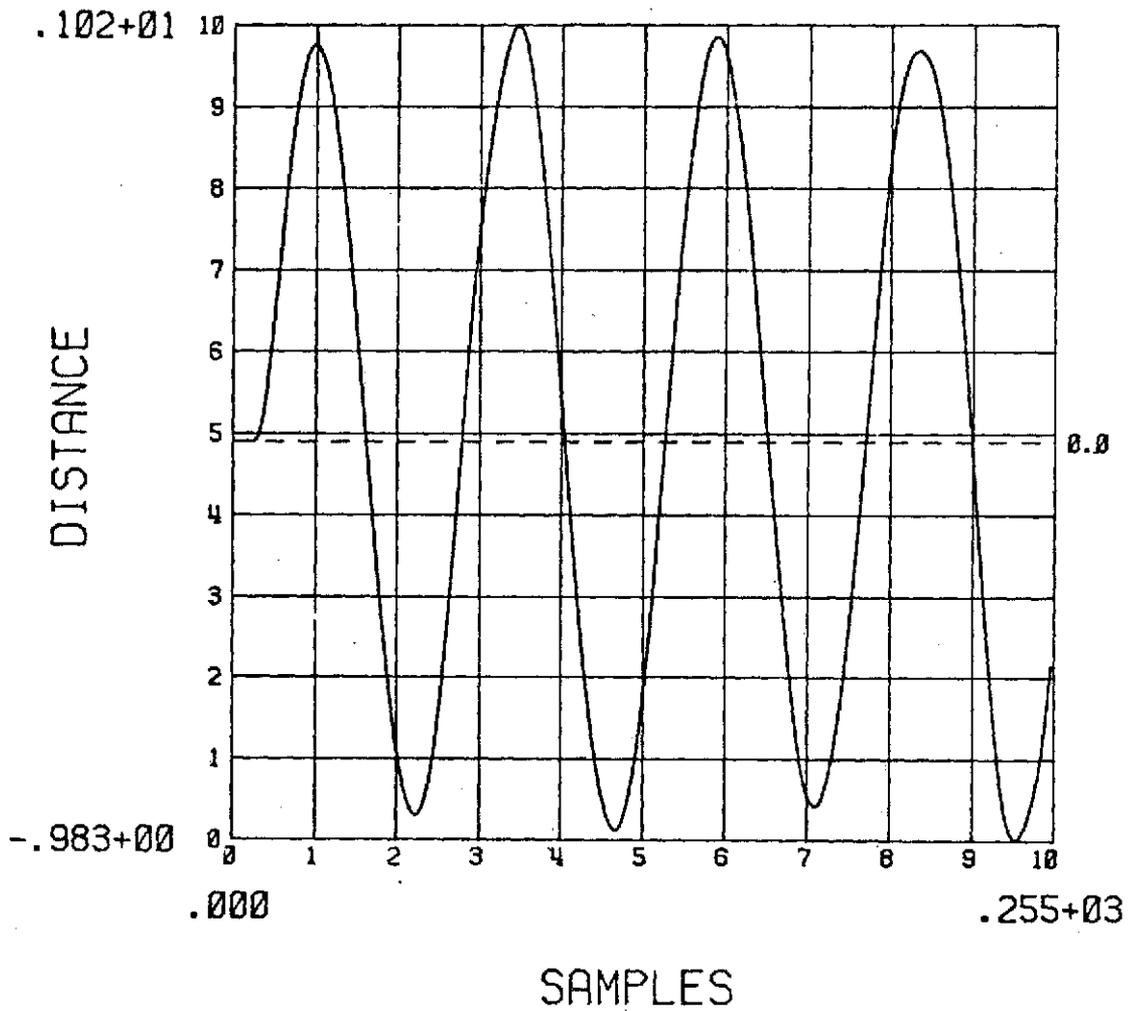


Fig. 4.20a. The response from DIRSIT to the third input. This is an estimate of the first derivative of the input signal (the first input).

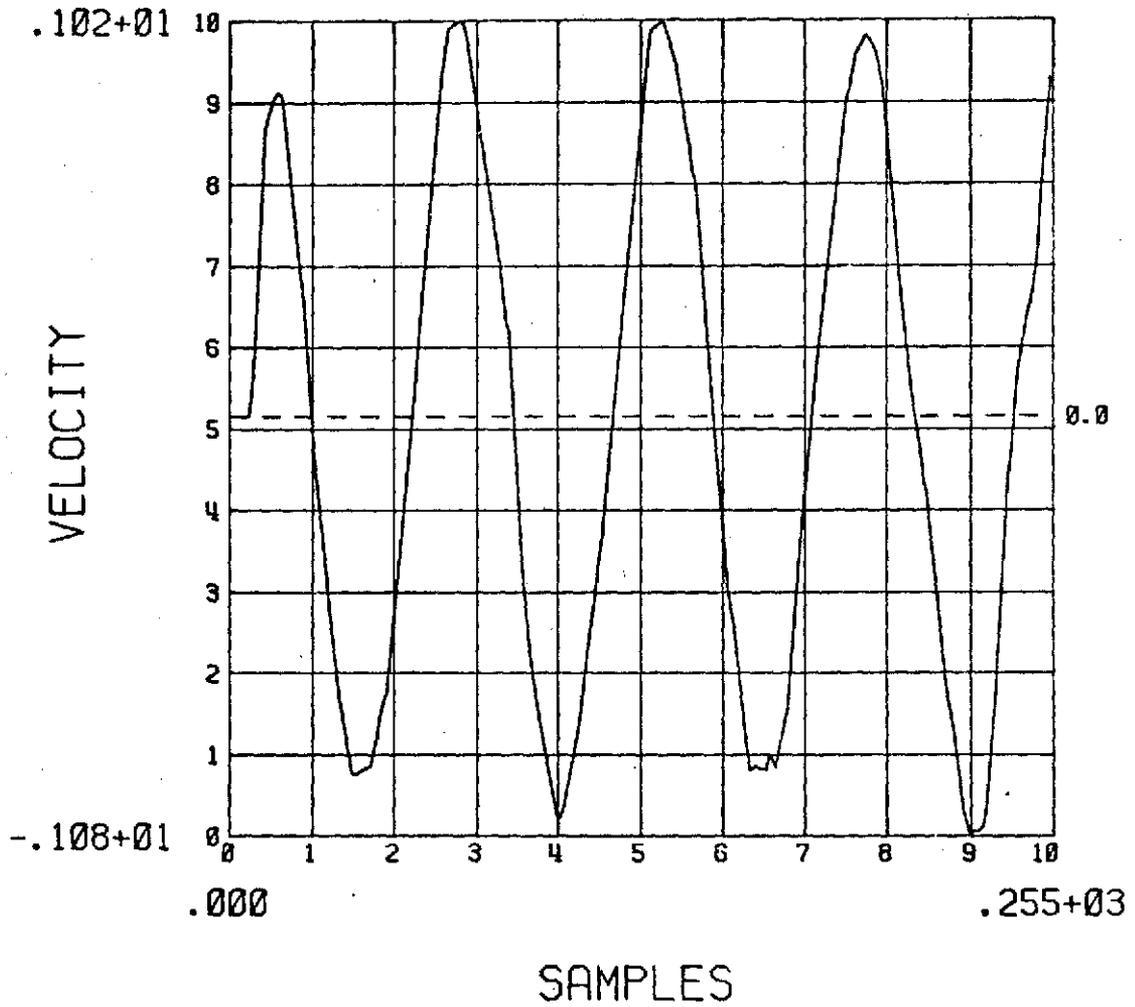


Fig. 4.20b. The response from DIRSIT to the third input. This is an estimate of the first derivative of the input signal (the first input).

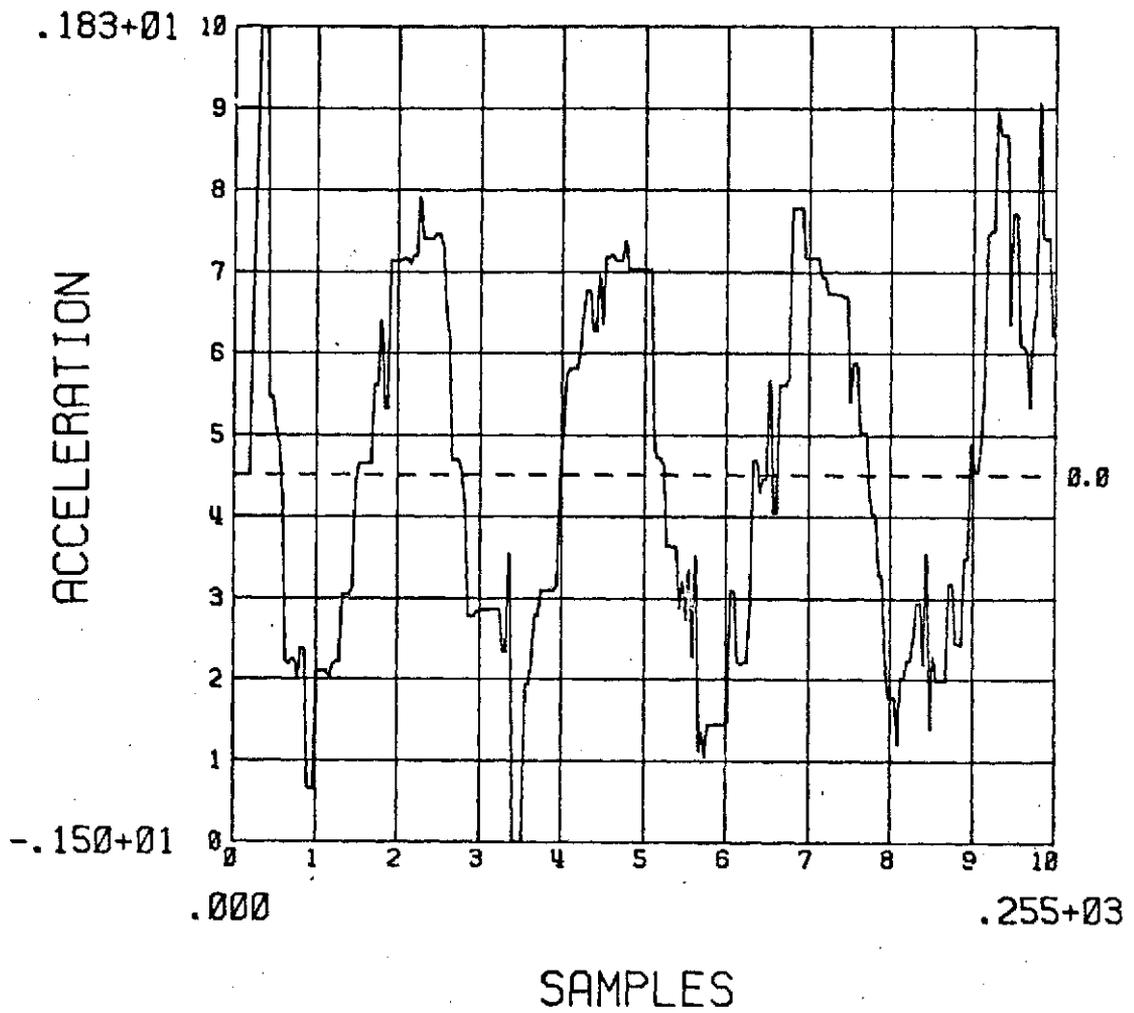


Fig. 4.20c. The response from DIRSIT to the third input. This is an estimate of the second derivative of the input signal (the first input).

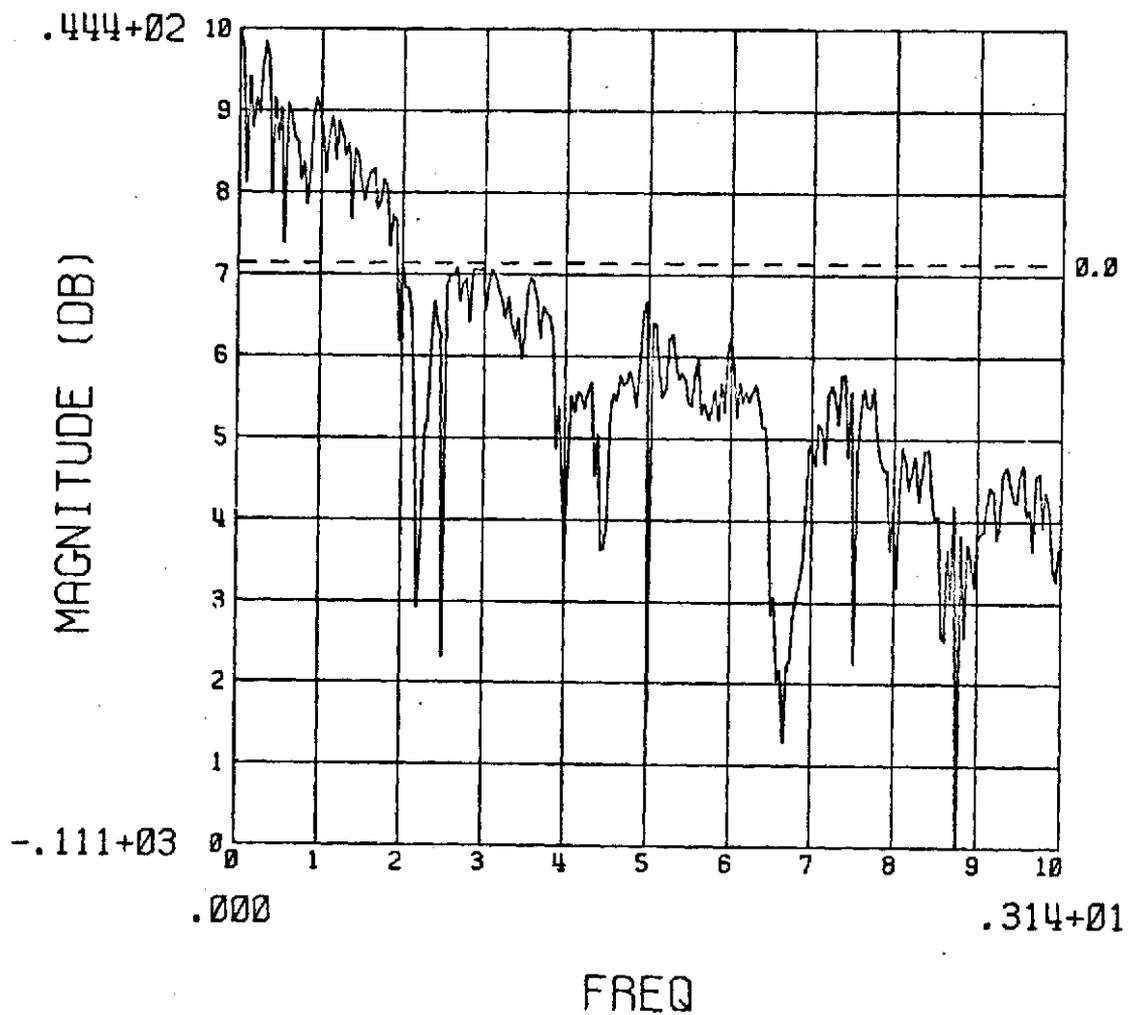


Fig. 4.21. The log magnitude of the response from the quasi-linear approximation to the third input (the frequency description).

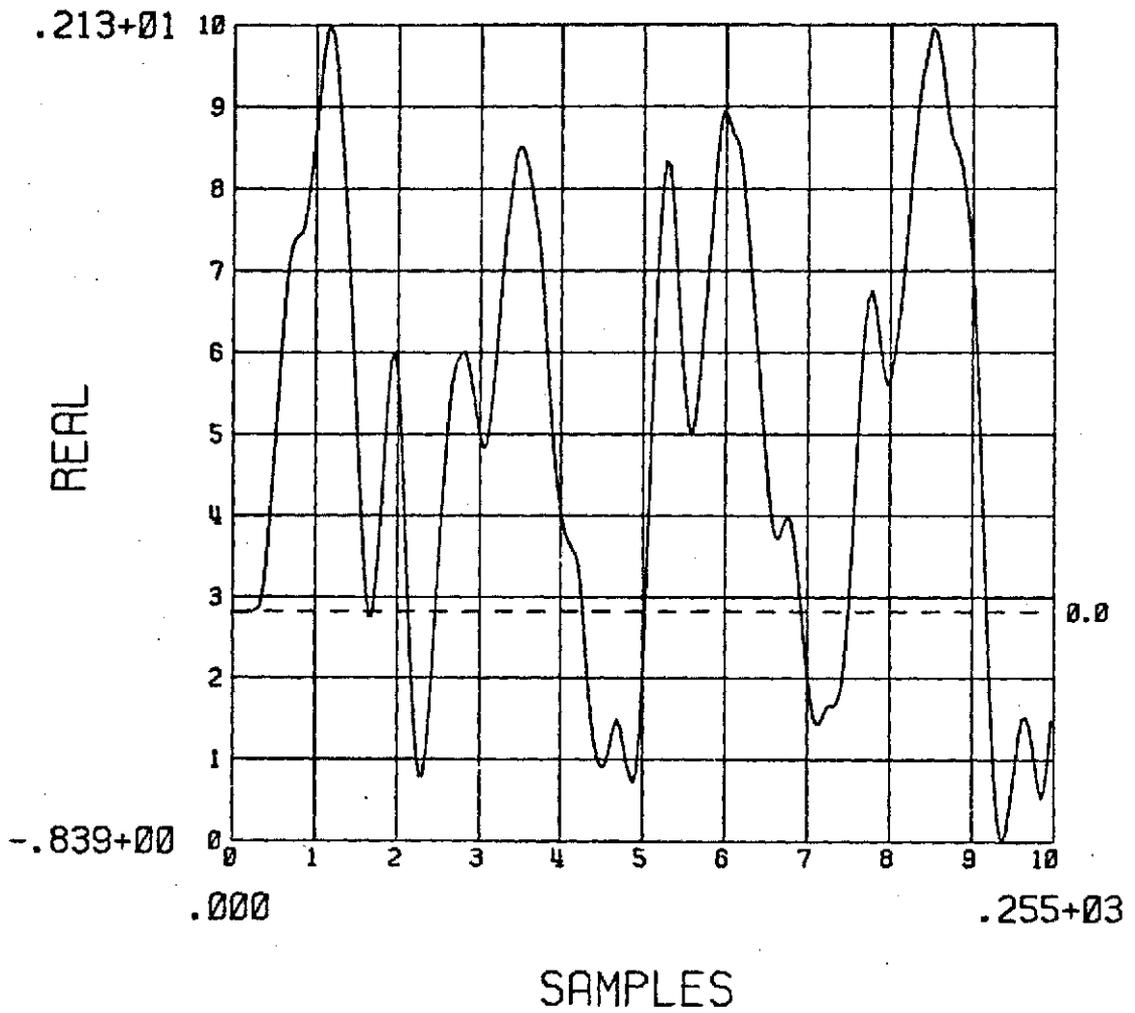


Fig. 4.22. The response from the quasi-linear approximation to the third input (the time description). Only the real part is shown, since the imaginary part is negligible.

estimate of DIRSIT. The quasi-linearization is seen to be very susceptible to the bias, whereas DIRSIT is immune. This is also illustrated in Fig. 4.23, which shows the difference between the DIRSIT output and the quasi-linearization output.

This latter input response difference is the main illustration of the power of the "nonlinear effect" of DIRSIT. It is the operation of DIRSIT on its nonlinear band that affords this immunity.

4.4 Comparison of DIRSIT to Other Systems

DIRSIT is compared with two other systems by feeding the three input signals of the last section through the other systems. This allows a qualitative comparison of DIRSIT as a discrete differentiator. The first system is a typical quadratic least-square fit to the data. This system yields the velocity and acceleration by differentiating the "best" quadratic curve fit. The second system is also a quadratic least-square fit to the data except that it uses the DST. The comparison of these systems with DIRSIT will yield a good idea of what benefits there are in the DST and what benefits there are in the way DIRSIT "fits" the data.

The first system is a typical differentiator that might be found in many program libraries. The differentiator windows the data with a window length of ten points. A second order polynomial is given a least-square fit to the data in the window and the coefficients are used to estimate the smooth curve and its first two derivatives. This system is referred to as the "quadratic least square" (QLS) system.

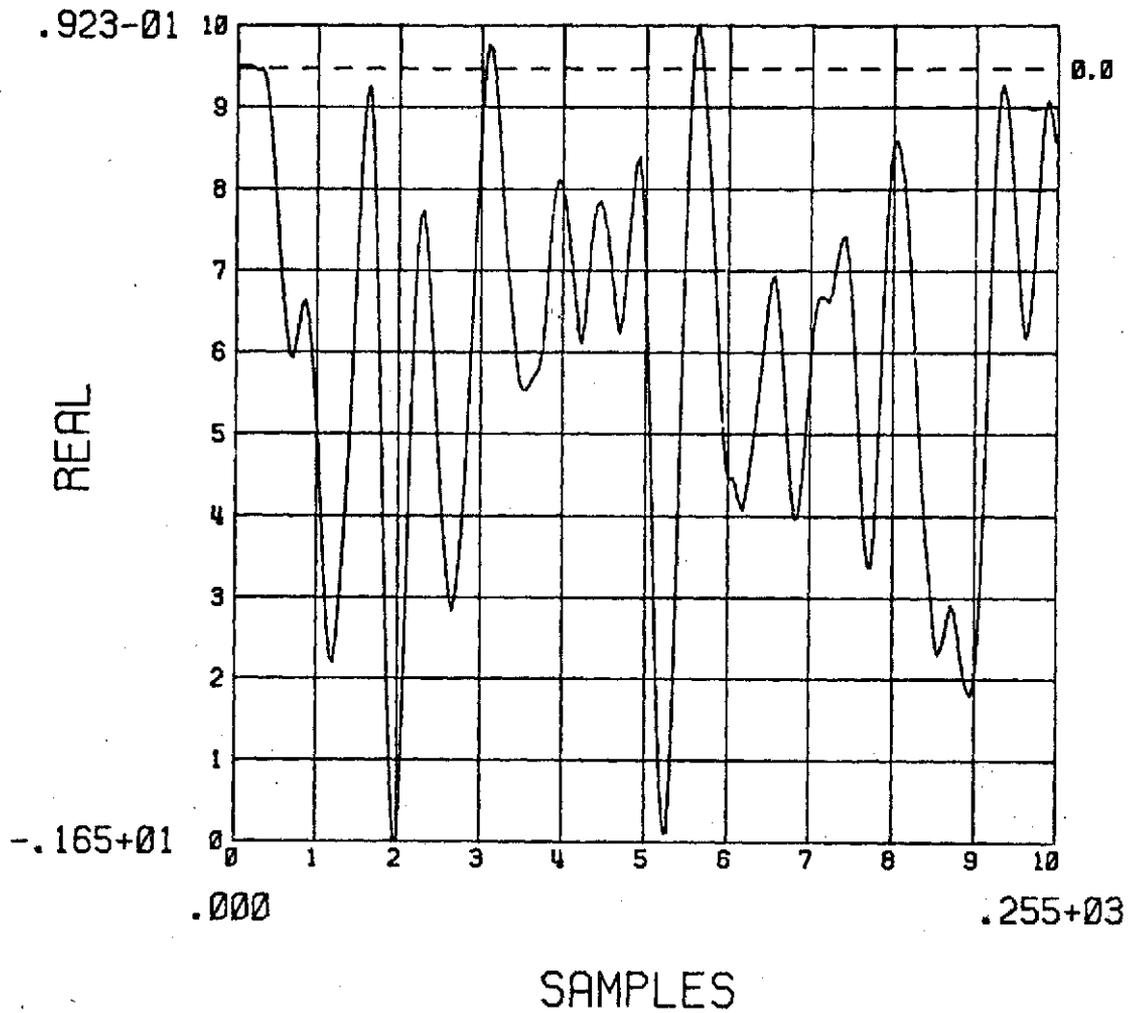


Fig. 4.23. The difference between the response from DIRSIT and the response from the quasi-linear approximation to the third input.

The QLS was chosen as an example to compare to DIRSIT for several reasons. First, the QLS is in common use. Second, in terms of the use of DIRSIT here at the University of Utah, the QLS is DIRSIT's main competition in the kind of data reduction DIRSIT is used for. And third, they both use the same model. They both assume that the third derivative of the "true" data is zero over the window.

The QLS system differs from DIRSIT in two important ways. First, the DST is not used; i.e., the smooth curve is fit to the data and not the acceleration. Thus, each new window is independent of the last one. Second, the "central tendency" is a least-square fit, which is much more susceptible to spikes than the median.

The output of the three inputs of the last section are shown in Fig. 4.24, Fig. 4.25, and Fig. 4.26, respectively. The output for the first input signal of this system and DIRSIT are rather comparable. The smooth estimate output for the second signal already shows effect from the noise. The velocity and acceleration get progressively worse. Finally, the output for the third signal is apparently complete nonsense.

This behavior is not completely surprising, since if a smooth signal has some contaminating noise in it, then its derivatives will have much more! The second system, denoted here as the quadratic least square using DST (QLSDST), may be expected to fair much better than the previous system on the first two inputs, since it gets away from differentiating noise.

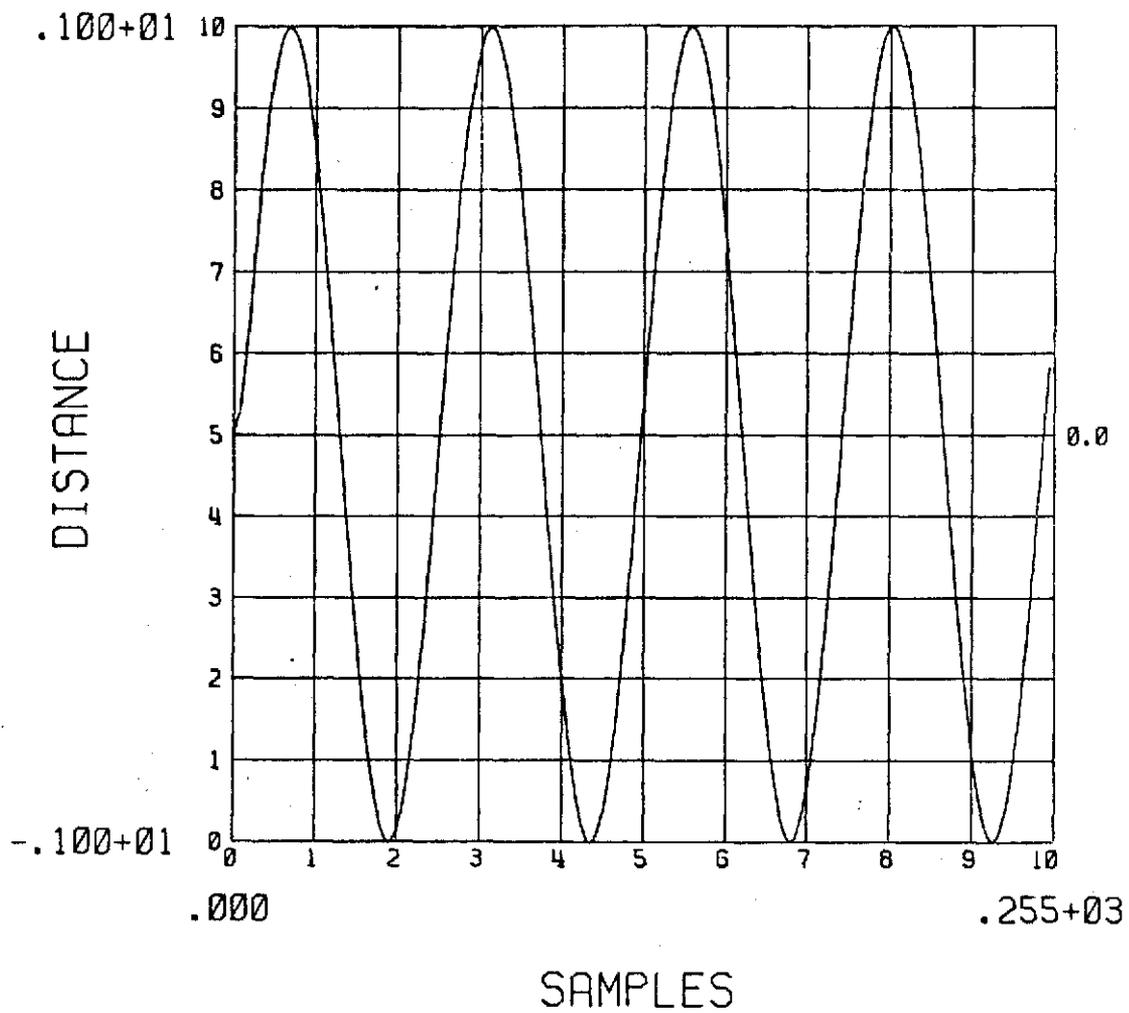


Fig. 4.24a. The response from QLS to the first input. This is an estimate of the input signal (the first input).

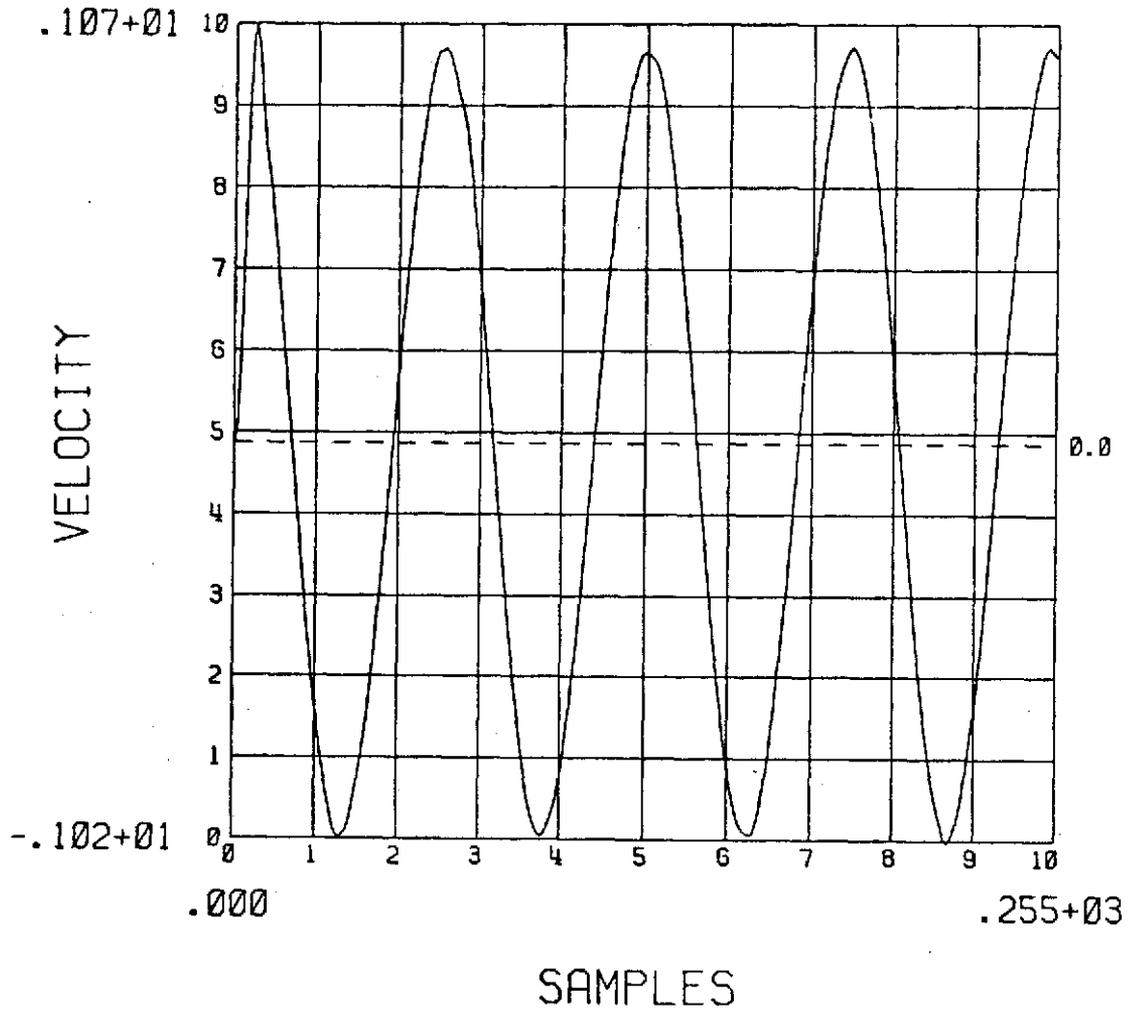


Fig. 4.24b. The response from QLS to the first input. This is an estimate of the first derivative of the input signal (the first input).

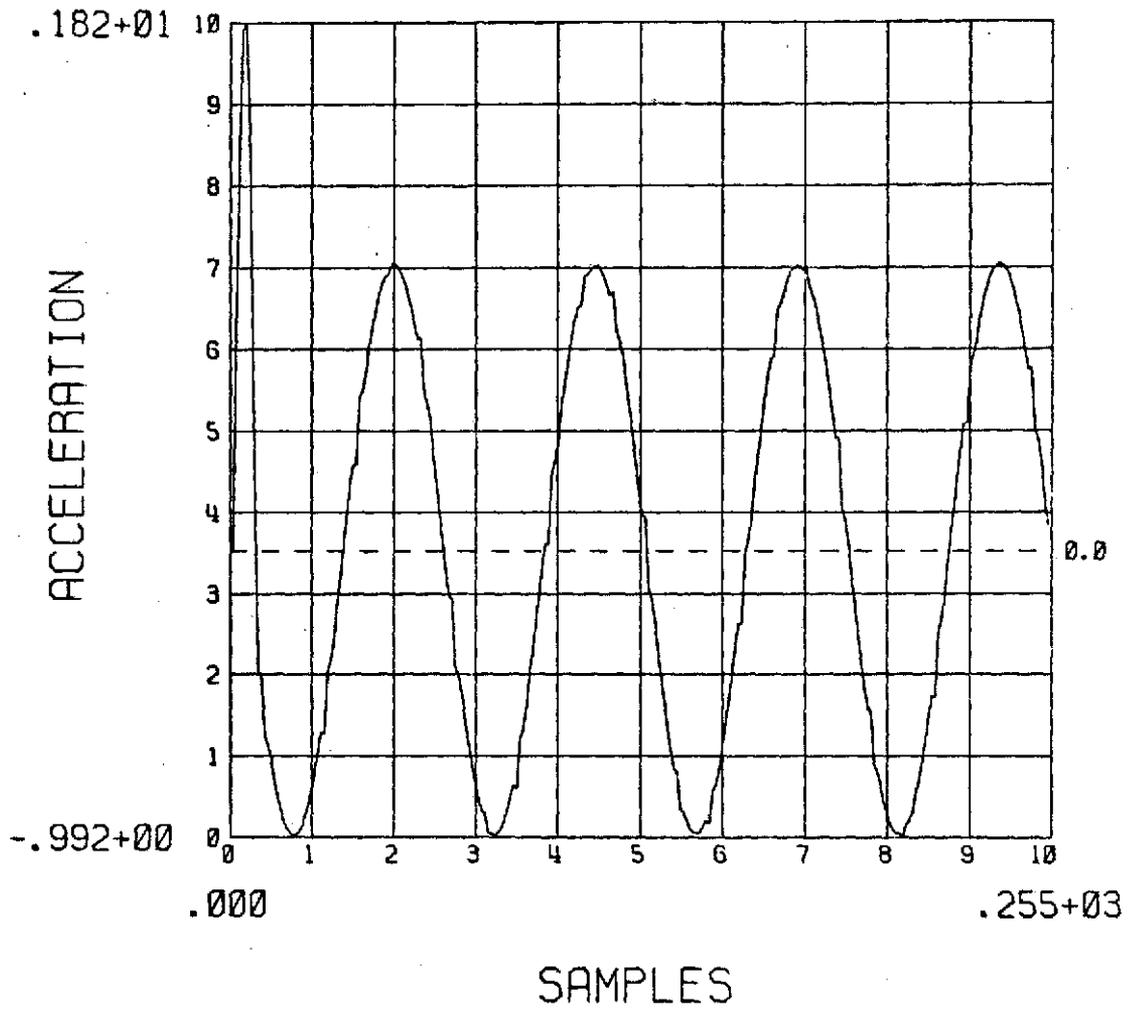


Fig. 4.24c. The response from QLS to the first input. This is an estimate of the second derivative of the input signal (the first input).

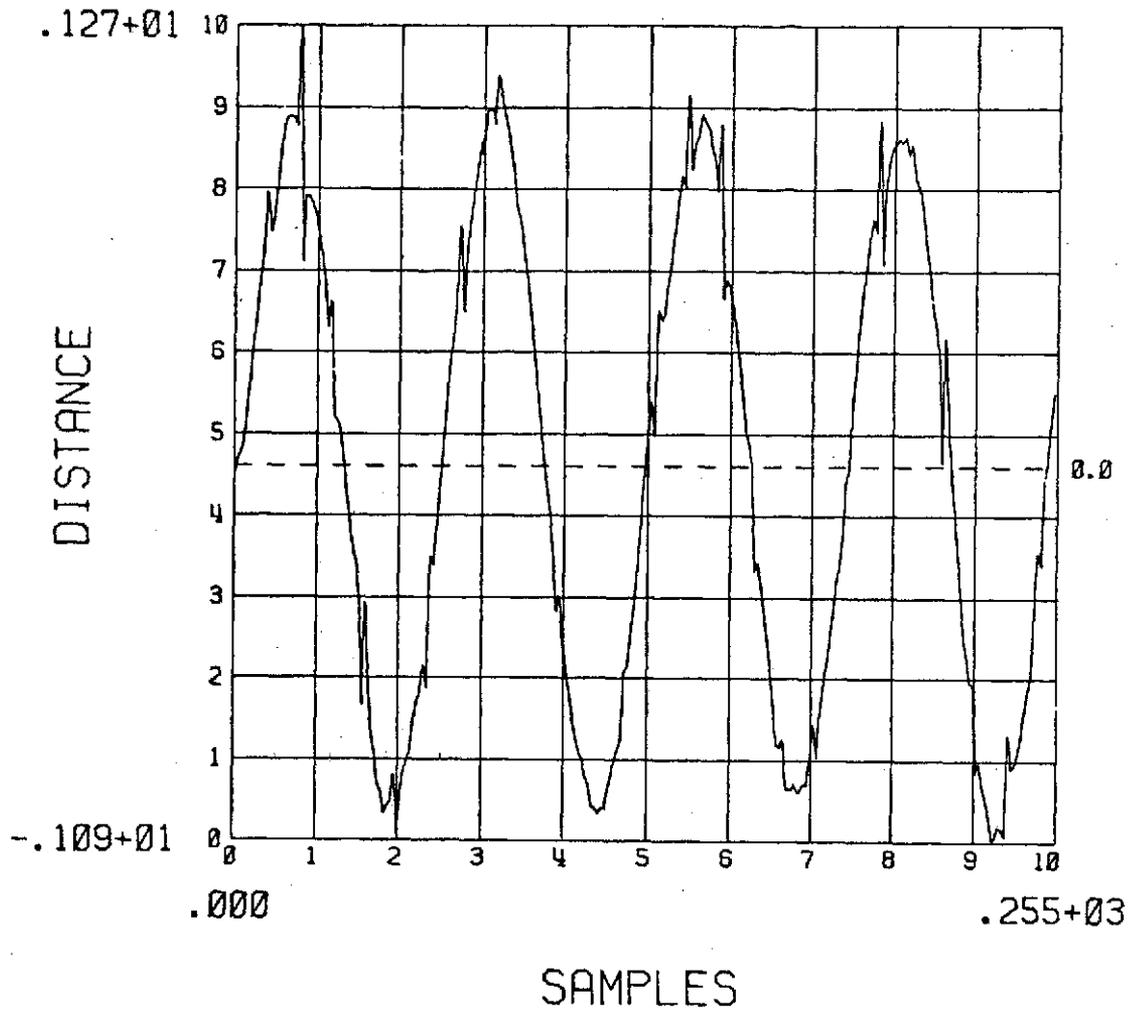


Fig. 4.25a. The response from QLS to the second input. This is an estimate of the input signal (the first input).

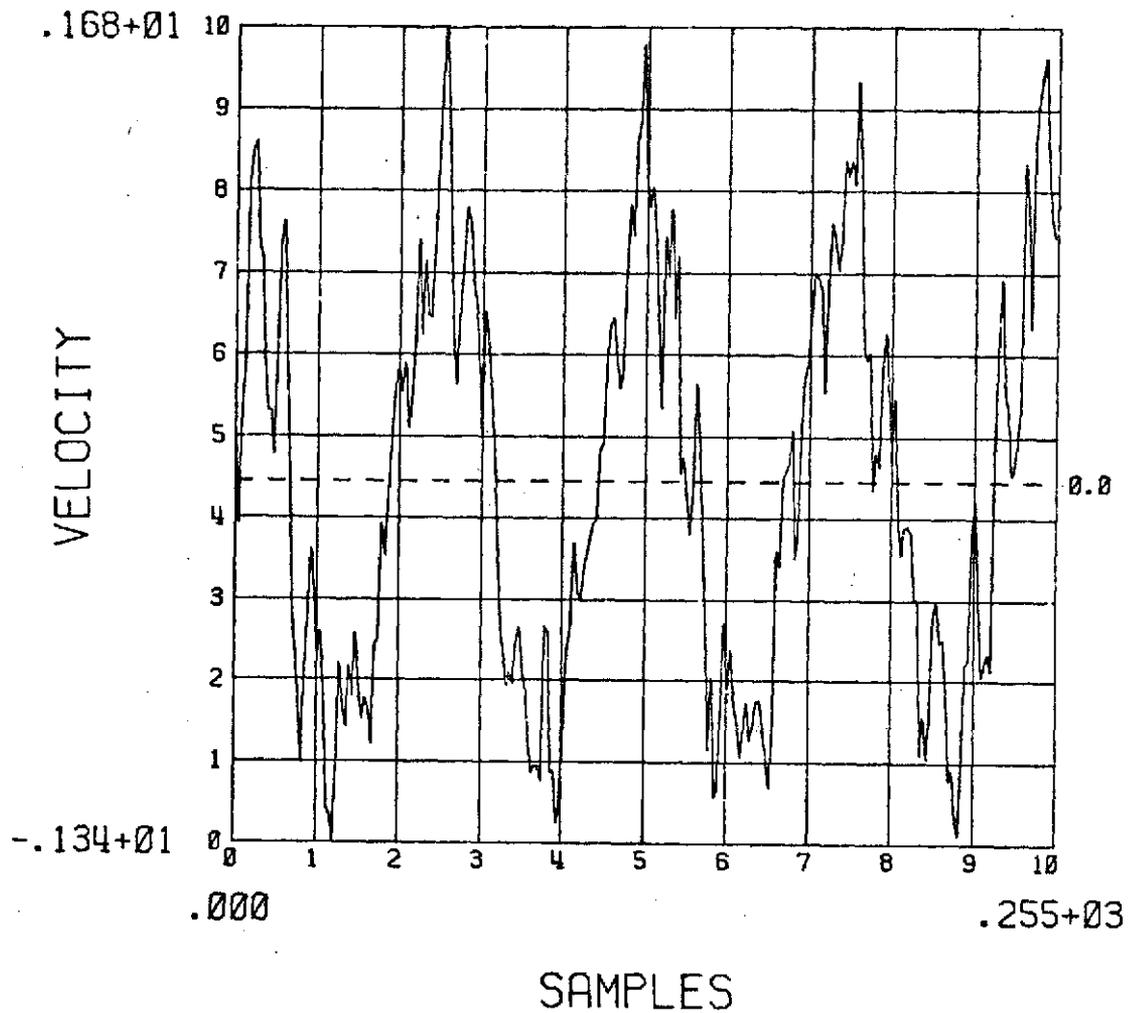


Fig. 4.25b. The response from QLS to the second input. This is an estimate of the first derivative of the input signal (the first input).

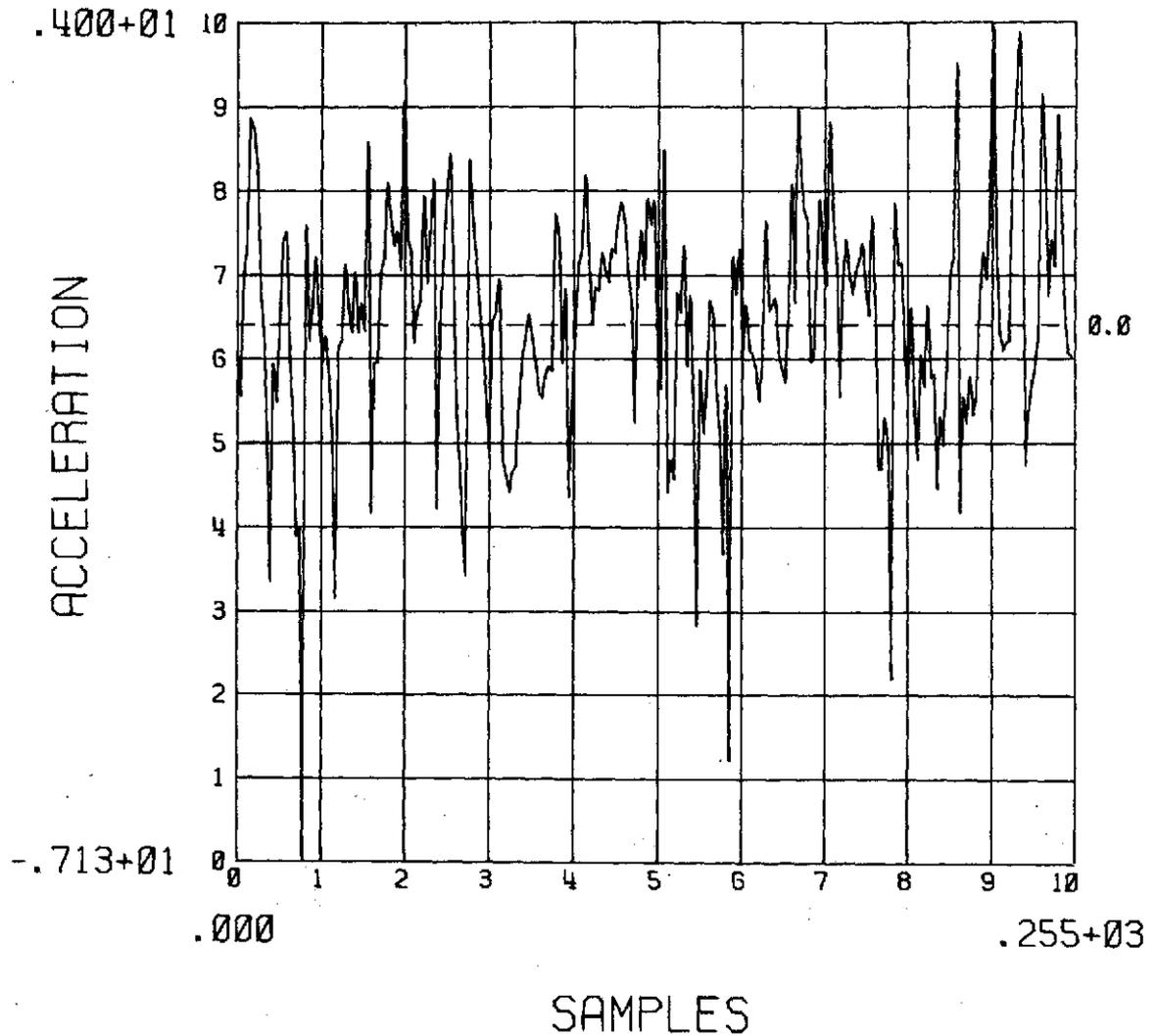


Fig. 4.25c. The response from QLS to the second input. This is an estimate of the second derivative of the input signal (the first input).

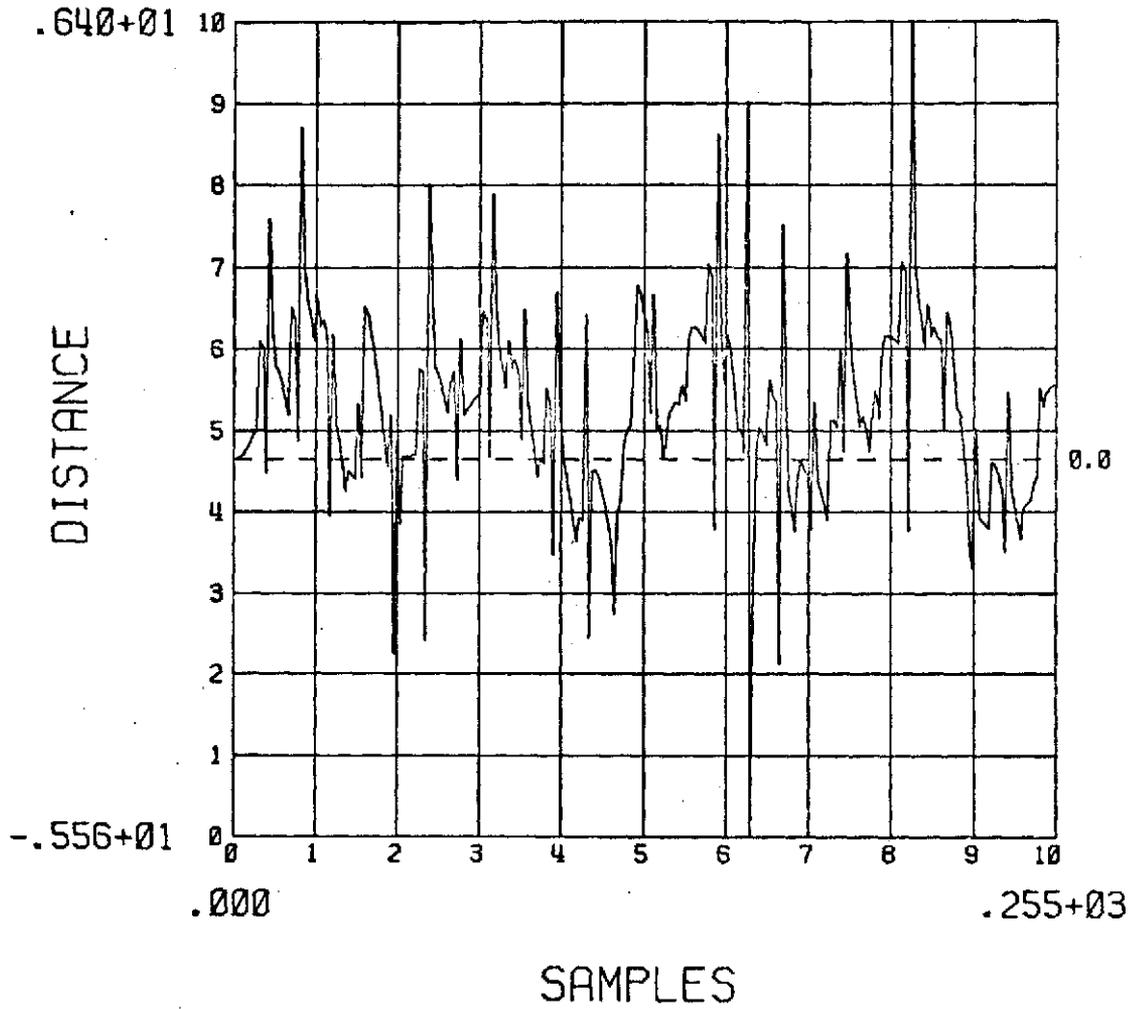


Fig. 4.26a. The response from QLS to the third input. This is an estimate of the first derivative of the input signal (the first input).

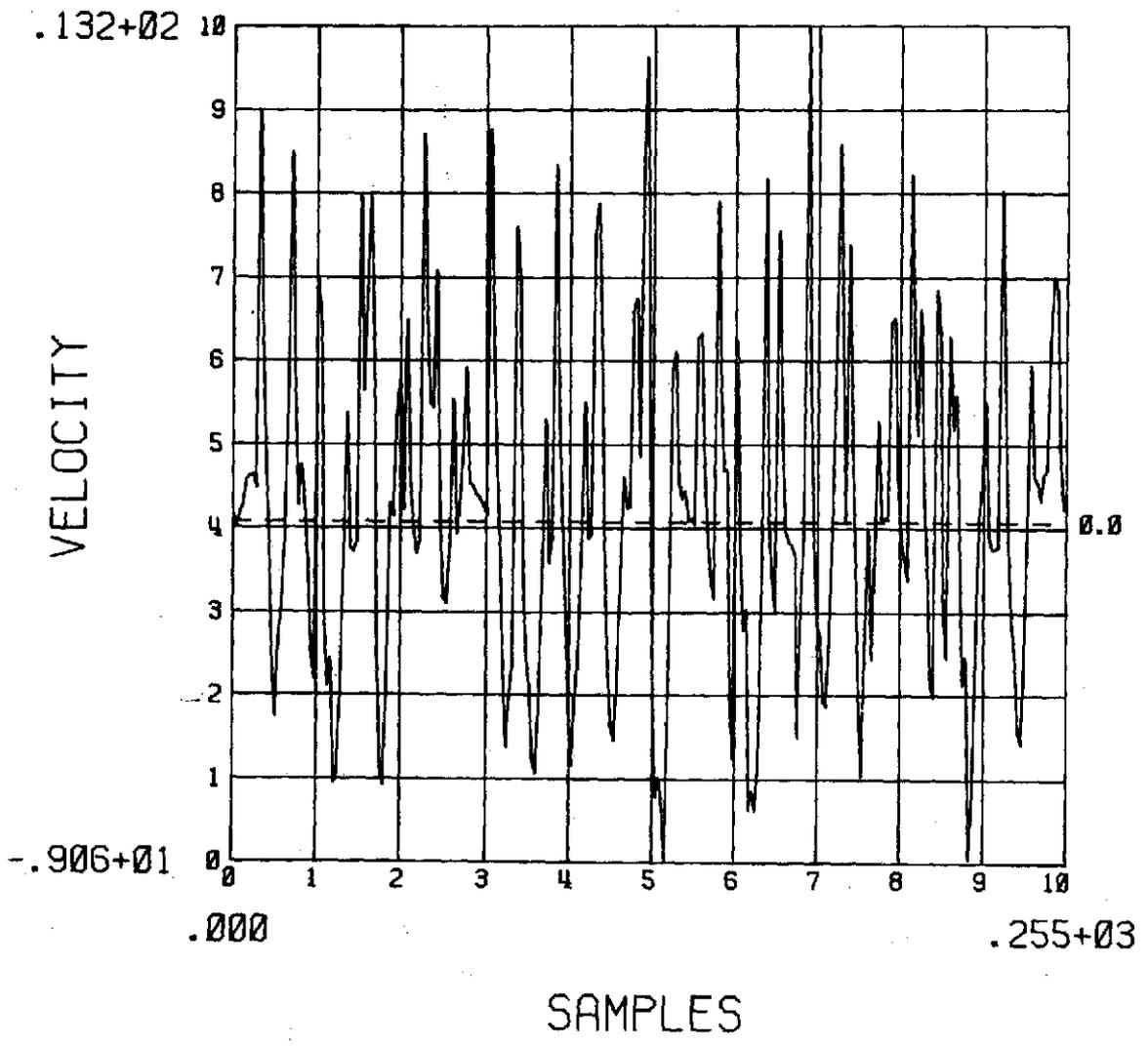


Fig. 4.26b. The response from QLS to the third input. This is an estimate of the first derivative of the input signal (the first input).

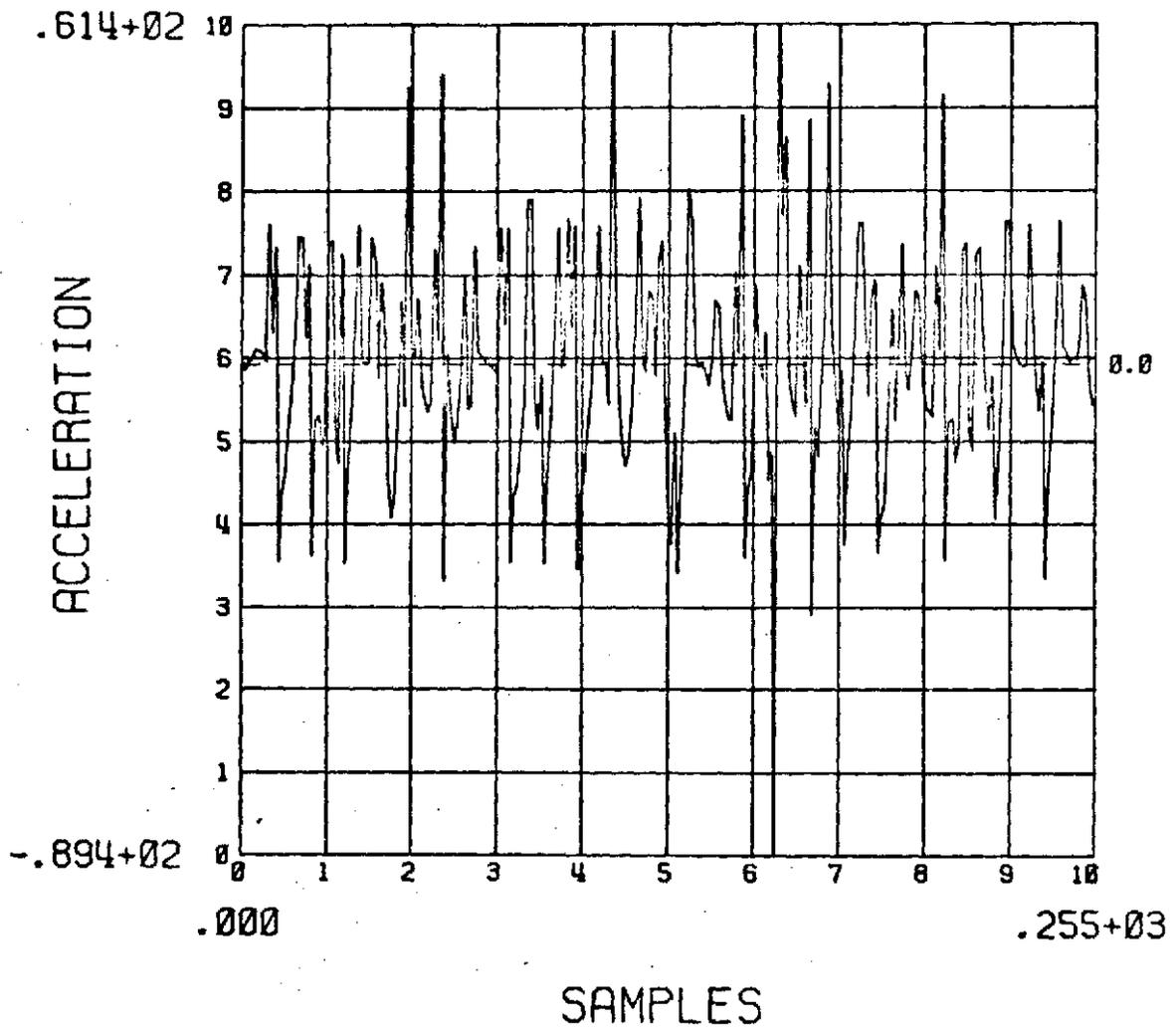


Fig. 4.26c. The response from QLS to the third input. This is an estimate of the second derivative of the input signal (the first input).

The output of the three inputs of the last section are shown in Fig. 4.27, Fig. 4.28, and Fig. 4.29, respectively. The outputs of the first two inputs are much better than for the previous system and rather comparable with DIRSIT. The output of the third input, however, shows the effect of the least-square "fitting" is much more susceptible to spikes than DIRSIT is.

These results characterize the two advantages of DIRSIT to QLS. As expected, the DST affords much better "filtering" than raw data filtering, since the differentiating of noise is avoided. Again, as expected, the median is seen to be much less susceptible to spikes than least squares.

4.5 Summary

DIRSIT has been shown to behave much like a low-pass filter for signals ranging from dc to 0.5 radians per sample. Above 0.5 radians per sample, DIRSIT exhibits nonlinear behavior. This nonlinear behavior effects a "spike immunity" and smoothing to signals in the linear region that have been contaminated with noise.

The "pass band" region comprises frequencies that have a half wavelength down to about the window length. Above this frequency, DIRSIT tends to reject and distort the data. Intuitively, this would seem to indicate that if the number of points in the window were changed, the corresponding "pass band" would change accordingly. This is indeed the situation. A DDF was found for a form of DIRSIT that utilized a window length of 6, and is plotted in Fig. 4.30. The ERF

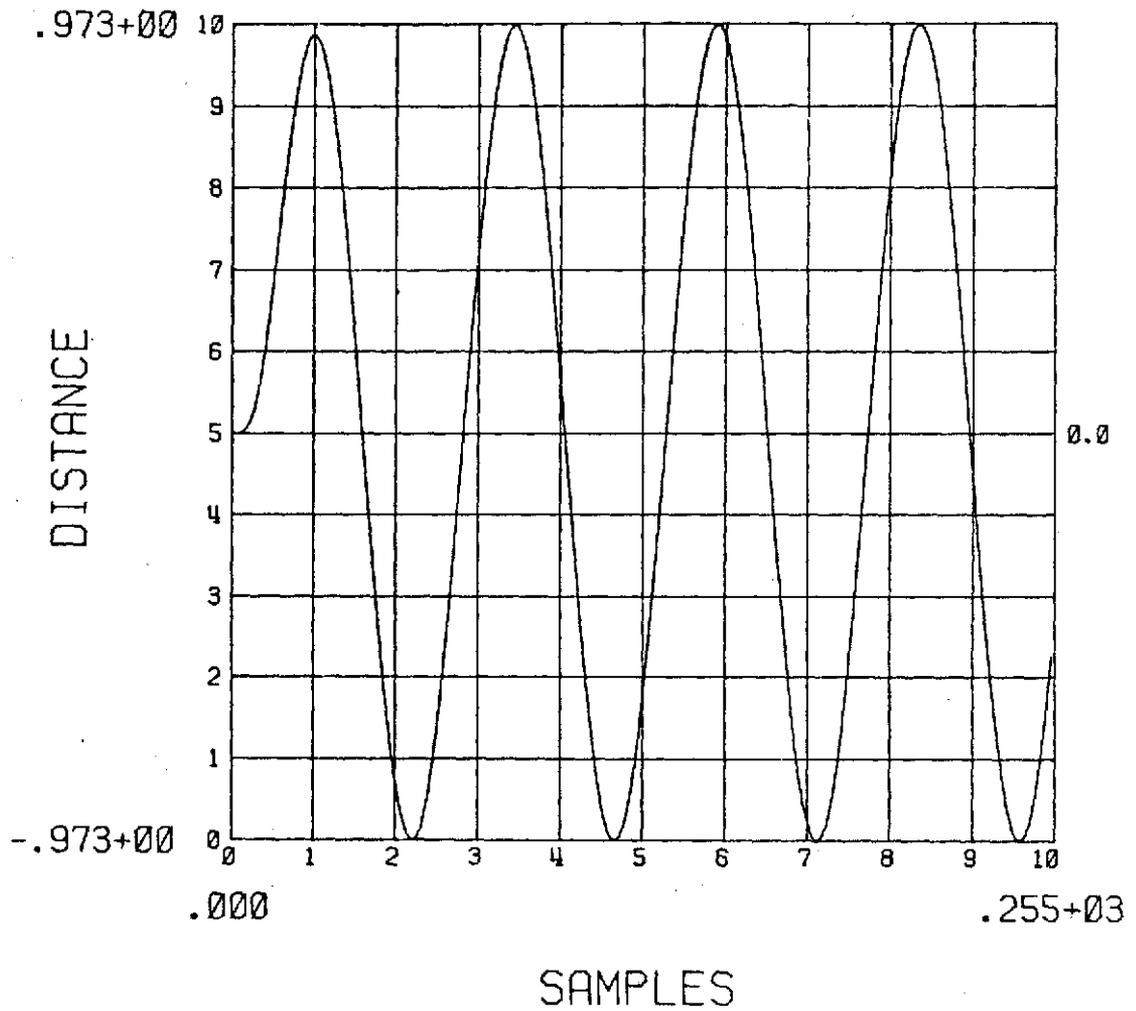


Fig. 4.27a. The response from QLSDST to the first input. This is an estimate of the input signal (the first input).

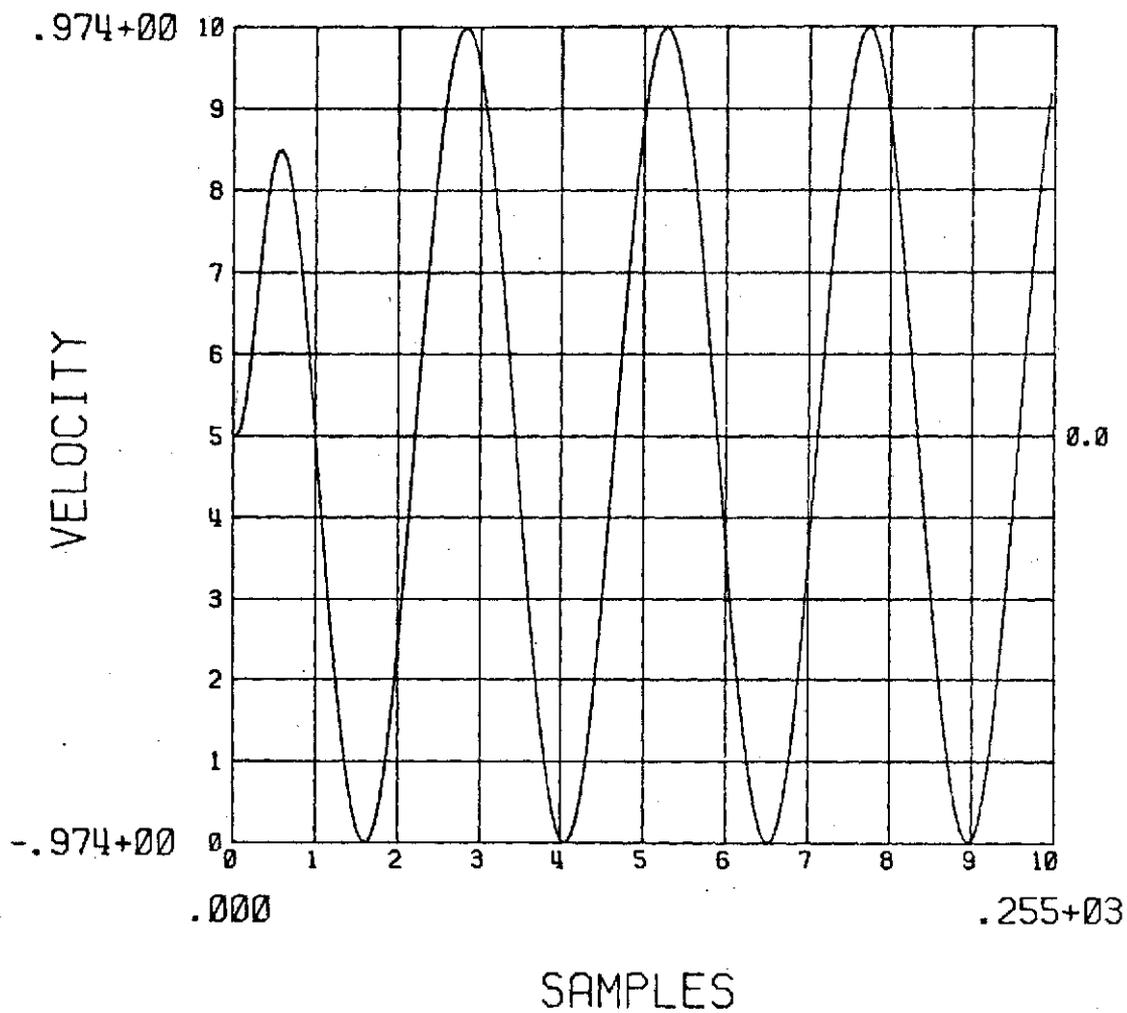


Fig. 4.27b. The response from QLSDST to the first input. This is an estimate of the first derivative of the input signal (the first input).

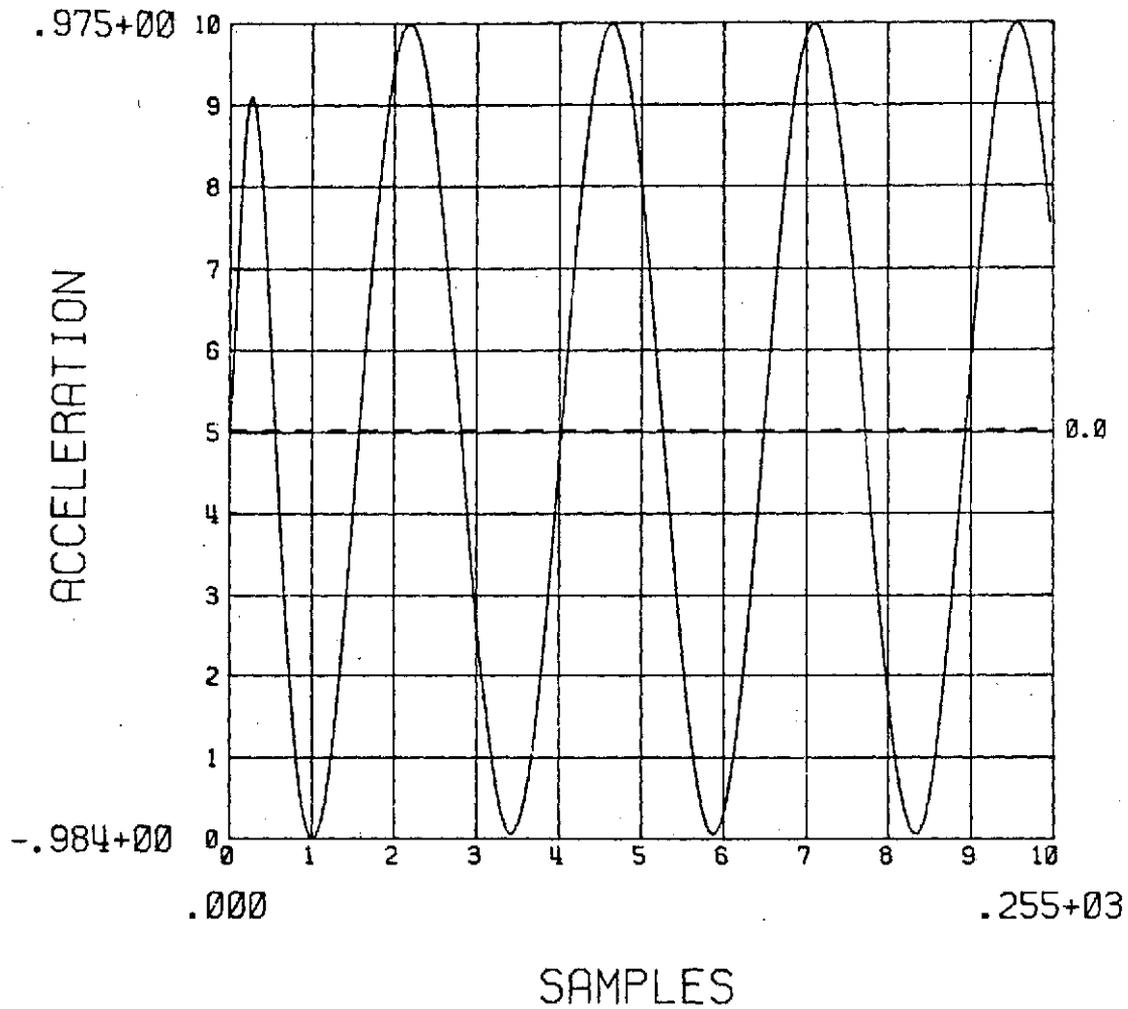


Fig. 4.27c. The response from QLSDST to the first input. This is an estimate of the second derivative of the input signal (the first input).

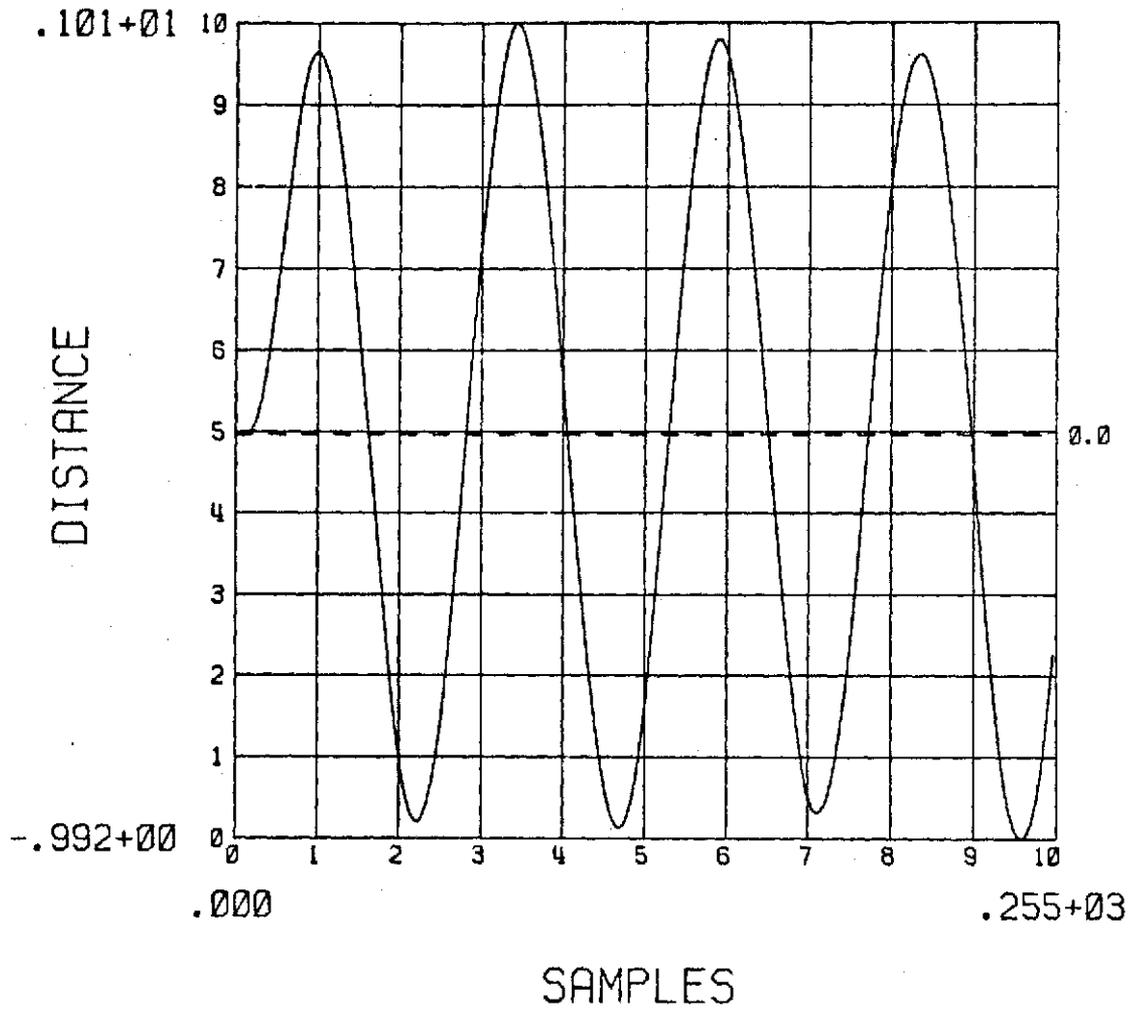


Fig. 4.28a. The response from QLSDST to the second input. This is an estimate of the input signal (the first input).

C-2

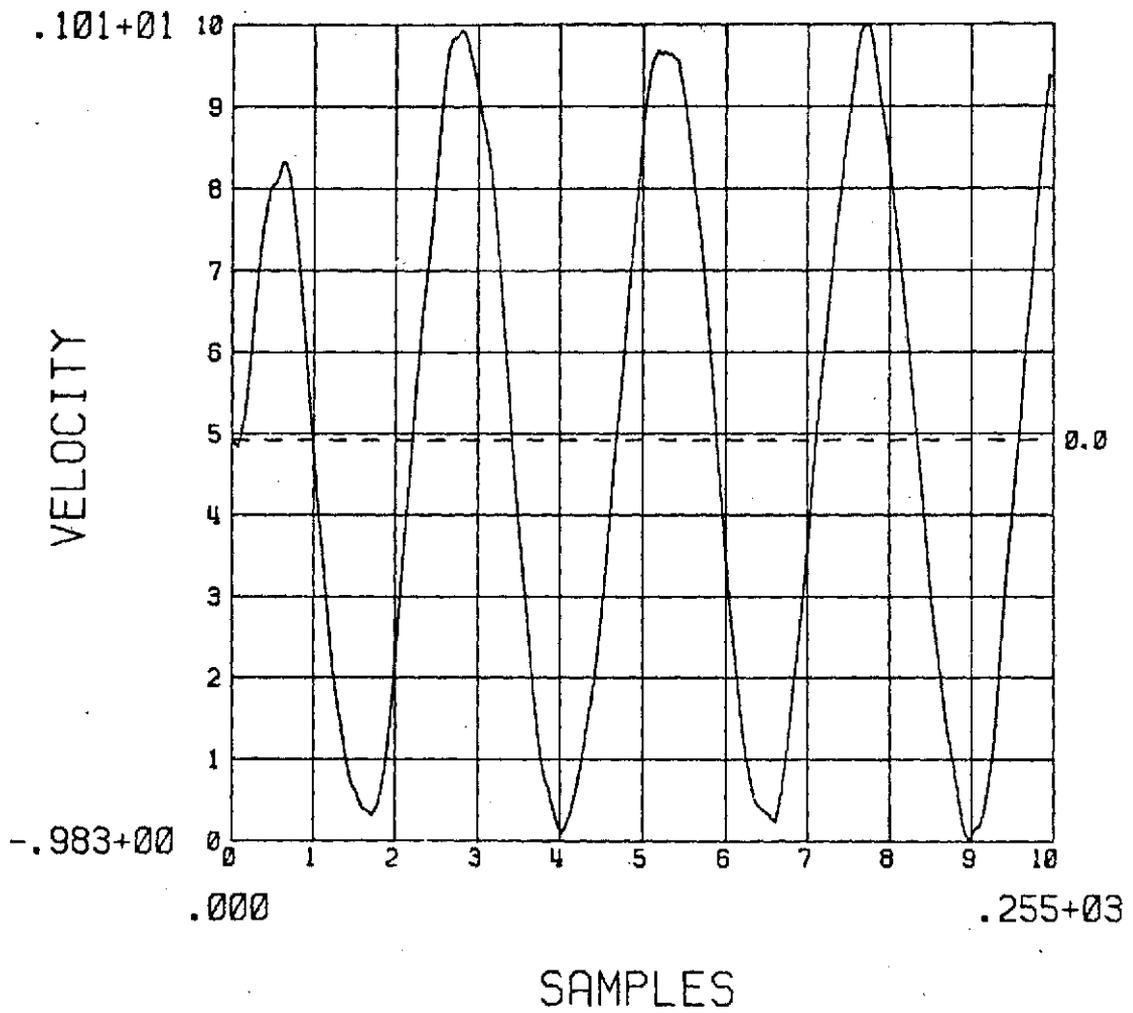


Fig. 4.28b. The response from QLSDST to the second input. This is an estimate of the first derivative of the input signal (the first input).

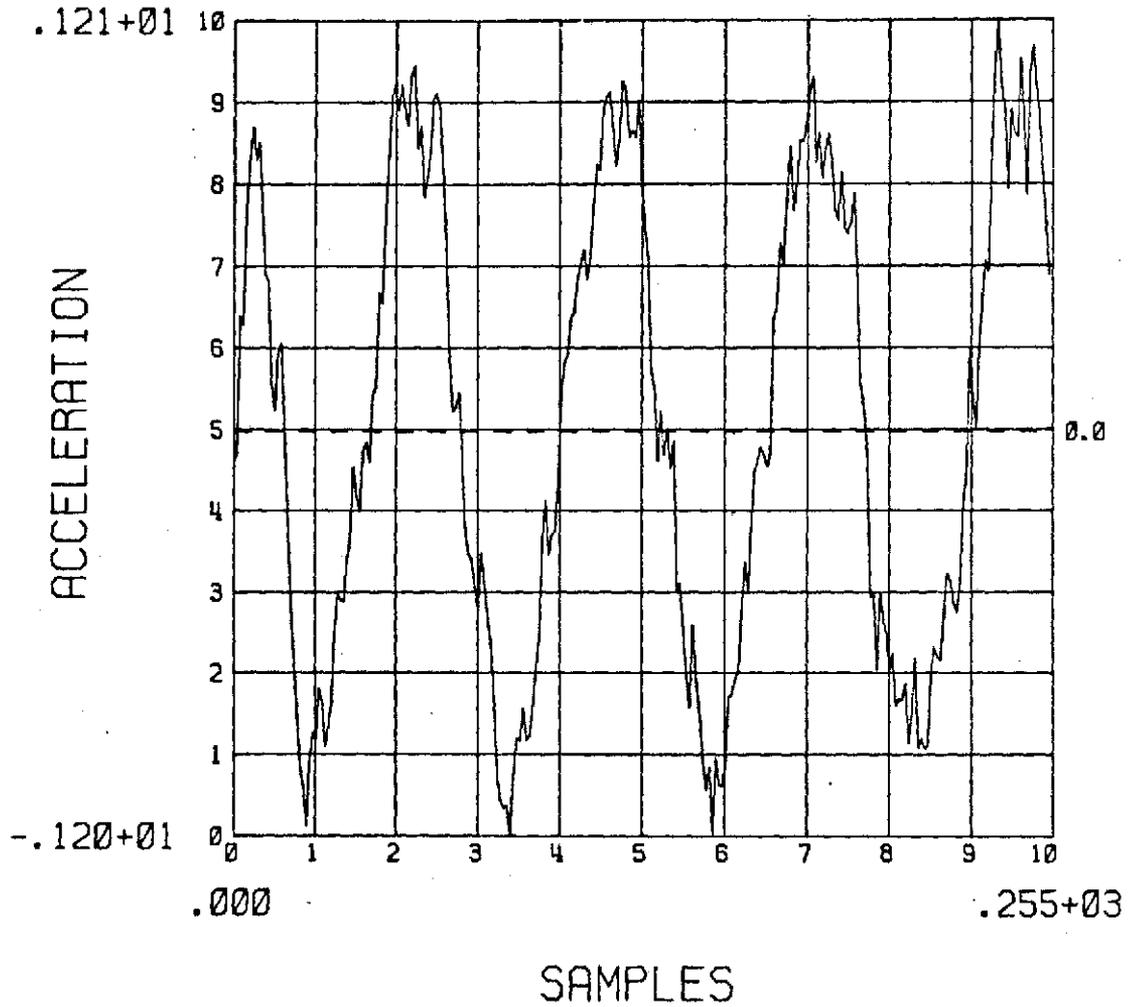


Fig. 4.28c. The response from QLSDST to the second input. This is an estimate of the second derivative of the input signal (the first input).

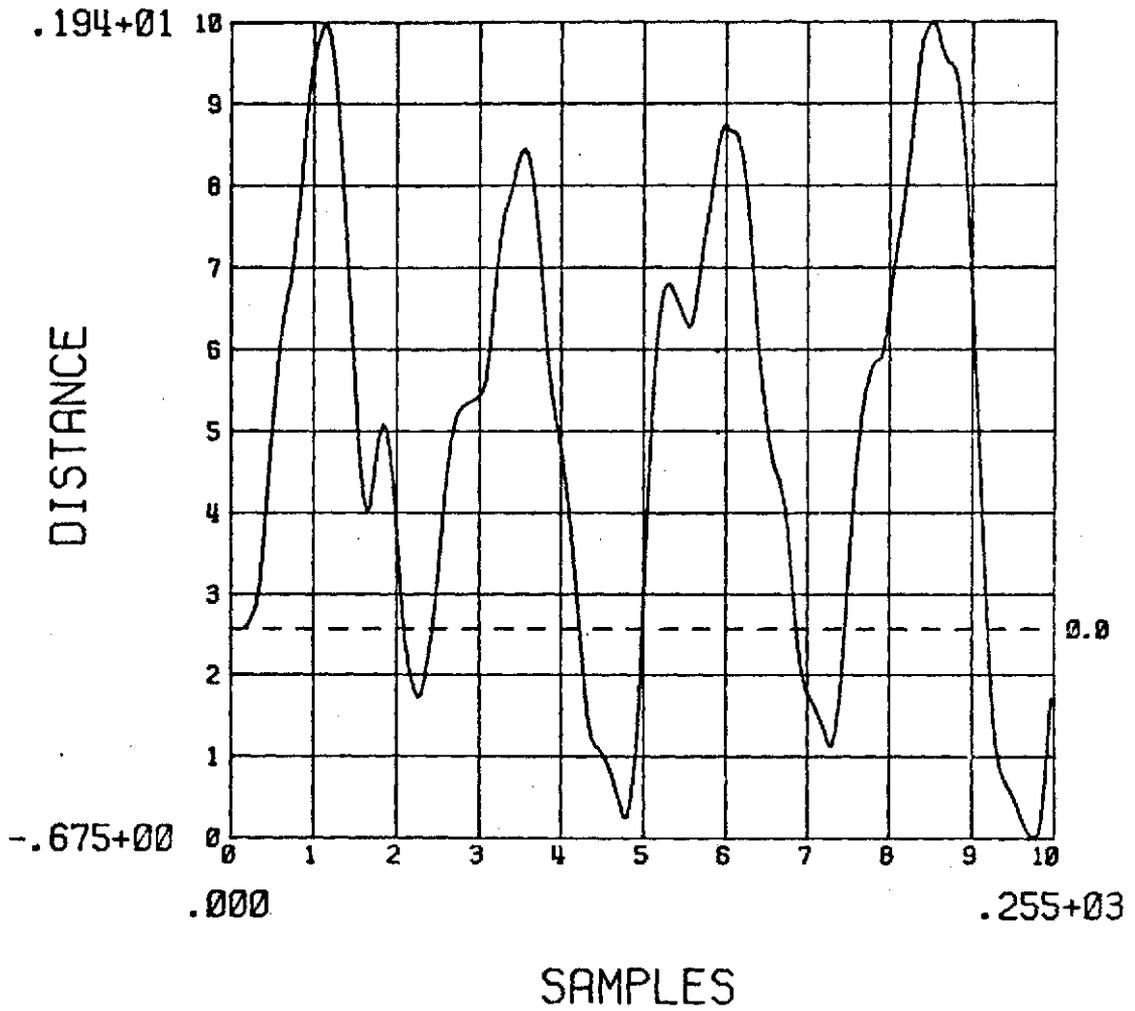


Fig. 4.29a. The response from QLS DST to the third input. This is an estimate of the first derivative of the input signal (the first input).

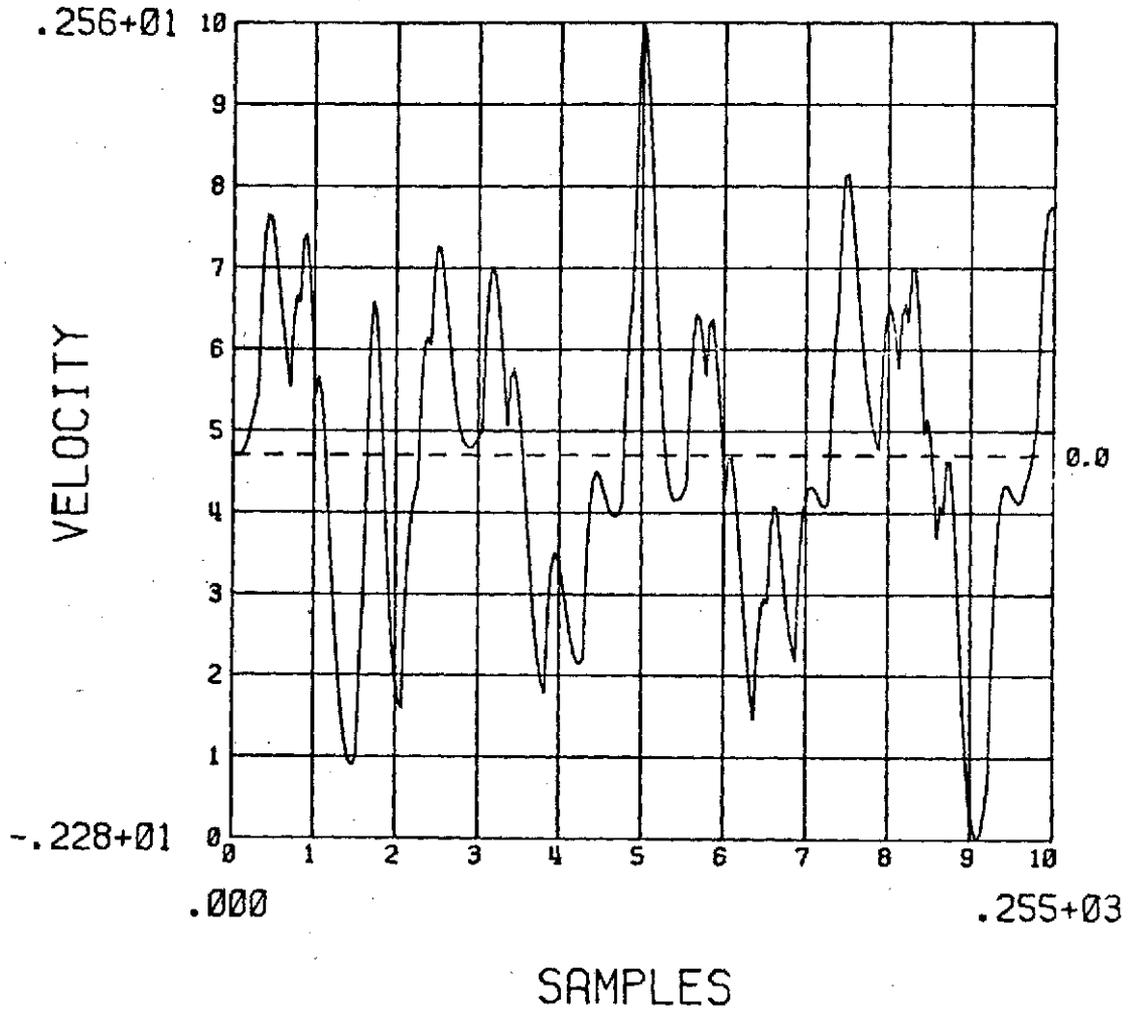


Fig. 4.29b. The response from QLSDST to the third input. This is an estimate of the first derivative of the input signal (the first input).

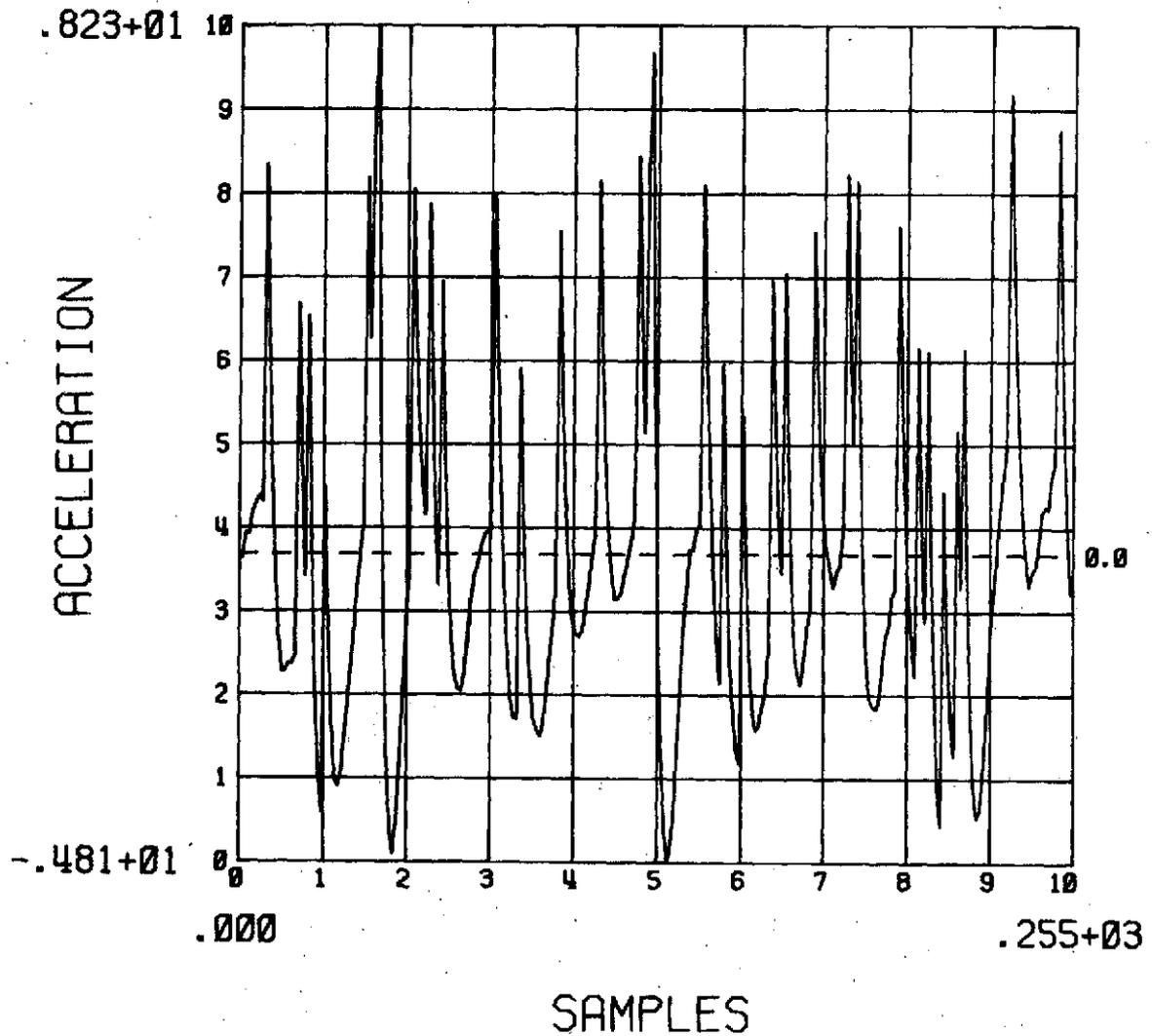


Fig. 4.29c. The response from QLSDST to the third input. This is an estimate of the second derivative of the input signal (the first input).

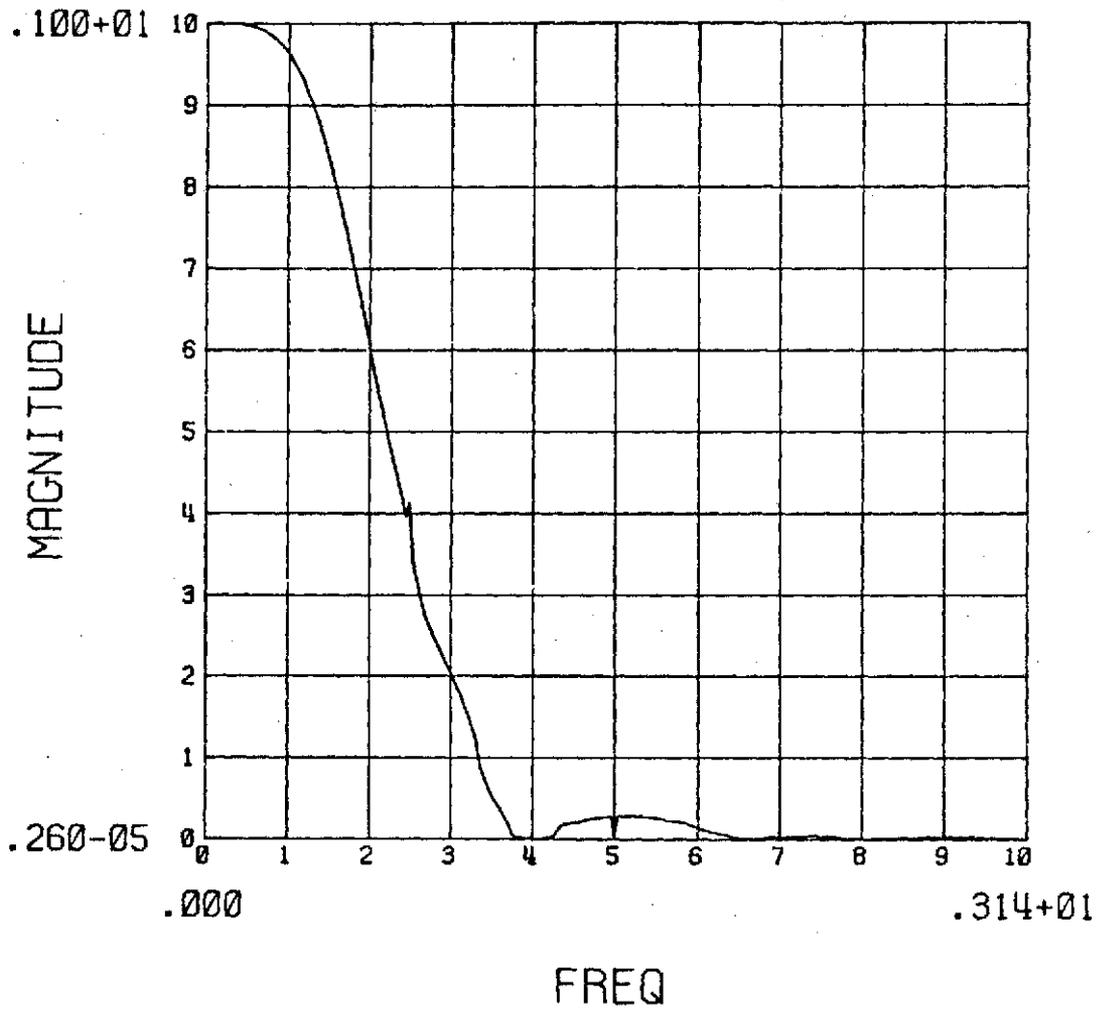


Fig. 4.30a. The measured DDF for the modified DIRSIT.
The magnitude (linear scale) is shown above.

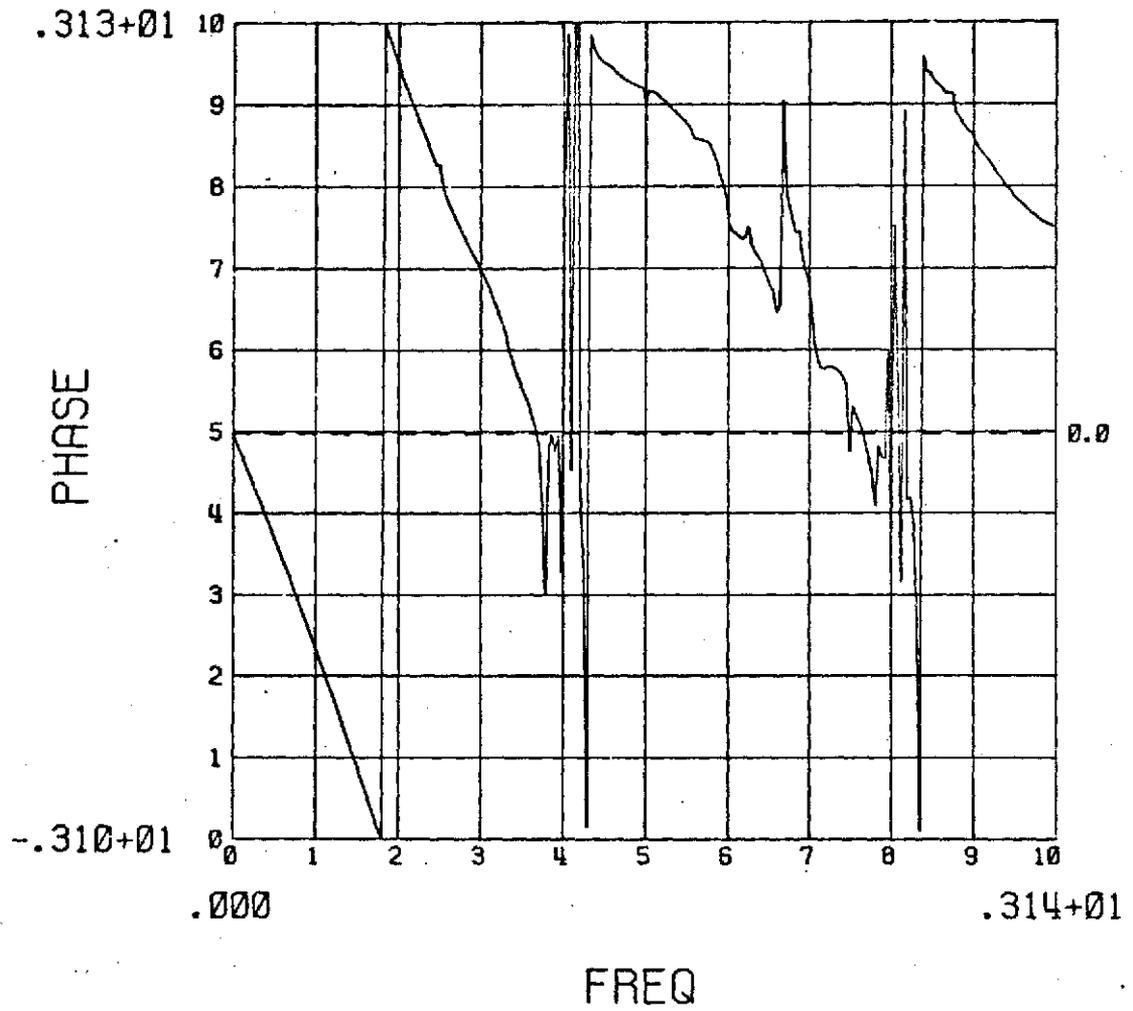


Fig. 4.30b. The measured DDF for the modified DIRSIT.
The magnitude (log scale) is shown above.

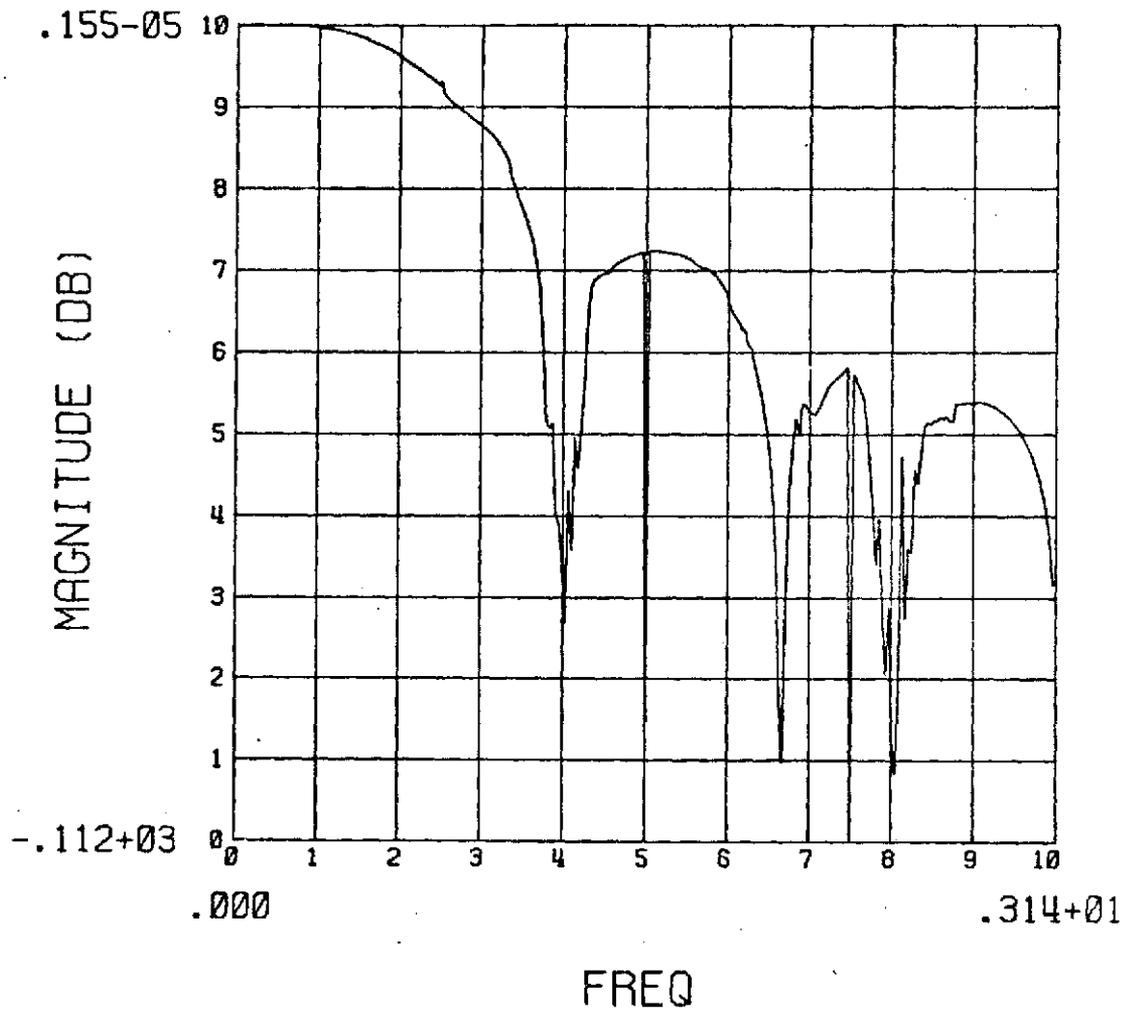


Fig. 4.30c. The measured DDF for the modified DIRSIT. The phase is displayed here.

is displayed in Fig. 4.31. These plots verify that the cutoff frequency for this version of DIRSIT IS about 12 samples per cycle, as predicted. This result adds a new dimension of flexibility to DIRSIT.

The model DIRSIT uses in its DST is that the second derivative is constant over the window. This amounts to a parabolic "fit". Other models could be used as well.

Further possibilities for investigation of DIRSIT using the present methods are:

1. A quantitative method could be implemented to "compare" digital differentiators.
2. Different models could be used to "fit" the data in a more appropriate manner.

This latter method would allow the tailoring of DIRSIT to a specific problem in much the same way that least-square approach is used now. Also, if the model allowed the evaluation of the third derivative, it is expected that this would yield much better estimates of the signal and the first two derivatives, since then the third derivative would "absorb" the noise.

The results, then, from this investigation show what to expect of DIRSIT. DIRSIT is seen to be immune to noise to a large extent by virtue of its use of DST, and immune to spikes by virtue of its use of the "median fit". By adjusting sampling and window length, DIRSIT exhibits a large degree of flexibility.

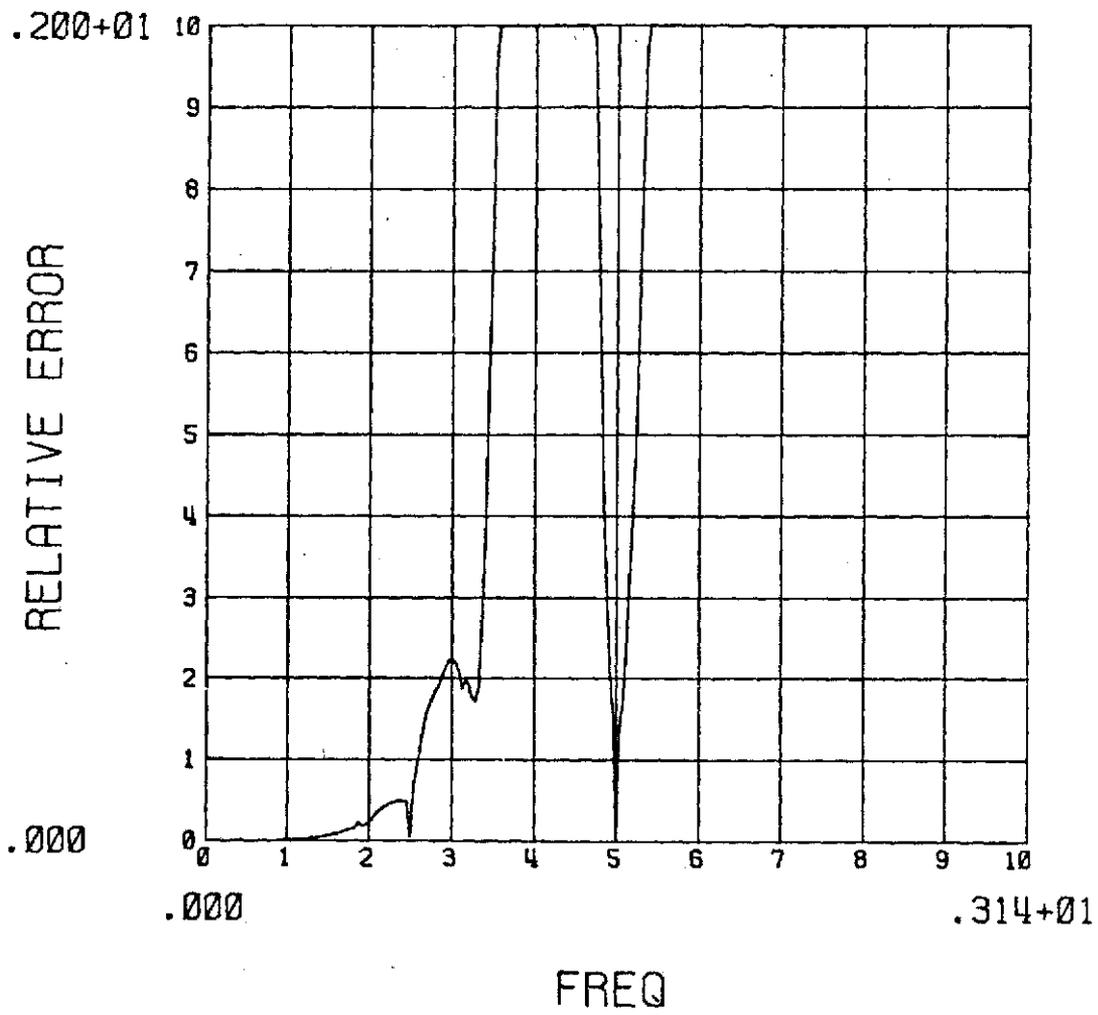


Fig. 4.31. The ERF for the DDF of the modified DIRSIT.

APPENDIX I

A VECTOR SPACE FOR DISCRETE SYSTEM ANALYSIS

In its most general sense, a physical system may be thought of as the realization of a mapping from one vector space to another. This is termed the "input-output" point of view. This approach is valuable since it allows the use of the mathematical tool box of functional analysis.

In the continuous case, both the input and output spaces may both be taken to be one of the L_2 spaces. Consider the case when the input is a sinusoid of period T , then assuming the output is periodic of period T , both spaces may be taken to be $L_2[0, T]$. An advantage to this approach is that $L_2[0, T]$ is a dense Hilbert space. That means that the function space may be spanned by a countable set of functions; i.e., any output may be expressed as an infinite series of any one of numerous countable bases sets. In particular, one very useful such basis set is the Fourier series, which allows all the familiar harmonic analysis techniques to be used. This Fourier theory implies that the response of a system to a sinusoidal input consists of the fundamental component (the best linear approximation) and the various harmonics. In general, a countably infinite number of harmonics are possible. This theory would, in turn, allow various filters to be used to "measure" the sinusoidal-describing function for an appropriate system.

The discrete case is more difficult. If a sinusoid is sampled in such a way that the ratio of the sampling frequency to the

sinusoidal frequency is rational, the input and output sequences would then be periodic of some period, say N . In this case, the input and output spaces could be taken to be the $\ell_2[1, N]$ space. As in the continuous case, $\ell_2[1, N]$ is a Hilbert space and Fourier theory applies. The basis set for this case could be $\exp(jk2\pi/N)$ for $k = 1, \dots, N$. The output of a discrete nonlinear system could contain only a finite number of harmonics, since $\ell_2[1, N]$ is finite dimensional.

If, however, the input sinusoidal is sampled in such a way that the ratio of the sampling frequency to the sinusoidal frequency is irrational, then the resulting sequence is not periodic. In this case, the input and output vector spaces must somehow contain all entries from negative infinity to positive infinity. The vector space used in this case cannot be a ℓ_2 space since it can contain no interesting vectors. In particular, the input sampled sinusoid is not contained in ℓ_2 . Some other space must be devised to allow the treatment of non-periodic, nontrivial sequences, preferably in such a way that for periodic sequences, the space reduces to an appropriate ℓ_2 space.

Consider the set A consisting of all complex sequences such that

$$\lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=-N}^N x(n) x(n)^* \quad (\text{A1.1})$$

is finite. The set A , together with the operations of addition and scalar multiplication as defined in the ℓ_2 space, is a vector space. The vector space becomes almost a pre-Hilbert space with the introduction of the inner product

$$\lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=-N}^N x(n) y(n)^* \quad (\text{AI.2})$$

for any x and y in A . The vector space V that is pre-Hilbert is a slight variation of A . V consists of almost all the vectors of A , except that if two vectors in A differ only at a few points, then they are considered the same vector in V . Formally, this amounts to the quotient space generated by the subspace consisting of all the vectors where Eq. AI.1 is zero. These concepts are examined more closely in Luenberger [3]. The important point is that if a comparison of two different vectors in V is made based on a comparison of two corresponding vectors in A , a region of the vectors in A must be selected such that they are representative of the vectors in V . The point here is that it is necessary to avoid regions that might contain finite duration transients.

An important property of the resulting space V is that it is not separable. To see this, it is only necessary to note that the two vectors $\exp(j\omega n)$ and $\exp(j\omega' n)$ are orthogonal, whenever ω and ω' are different. But there are an uncountable number of these vectors, one for each real number ω . Thus, there is an uncountable set of mutually orthogonal vectors contained in V , so that no countable basis set could span the entire space. The significance of this result from the engineering standpoint is that there could very possibly be an output of a system that was orthogonal to every term of some appropriate Fourier-type expansion, so that no analysis from that standpoint is possible!

Another important result is that, for the case where the sequences are periodic of period N , the vector space V is isomorphic to $\ell_2[1, N]$. This may be seen by noting that

$$\begin{aligned}
 (x, y)_V &= \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{n=-N}^N x(n) y(n)^* \\
 &= \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{k=-M}^M \frac{1}{N} \sum_{n=kN}^{kN+N-1} x(n) y(n)^* \\
 &= \frac{1}{N} \sum_{n=kN}^{kN+N-1} x(n) y(n)^* = (x, y)_{\ell_2[1, N]}
 \end{aligned}$$

where it is assumed that $n = kN, \dots, kN + N - 1$ is a "representative region".

Finally, V contains the vectors of interest. Any sampled periodic function is in V , as well as many sampled nonperiodic functions. The vector space V , as has here been briefly presented, is useful to handle nonperiodic as well as periodic discrete input-output system relations.

APPENDIX II

THE 5-T'S METHOD

The 5-T's method is utilized to realize a finite impulse response (FIR) approximation to a discrete system that has a specific frequency specification. The method consists of guessing the FIR of the system in an intelligent way.

The problem may be stated as follows. Given samples of the frequency response desired, find a FIR with the desired frequency response to within some specified tolerance to realize an approximation of the desired system. In some cases, the problem may be to realize a FIR approximation to a filter specified by a continuous frequency description. In fact, a practical filter cannot be given a continuous distribution in general, since it would require an ink of width zero! Thus, in all but "textbook" examples, the frequency description is given by discrete samples. The case considered here is when these samples are evenly spaced from 0 to 2π . The problem is to guess an appropriate FIR.

Whatever the "real" system is, it has a specific unique impulse response $h(n)$. In almost all practical cases, $h(n)$ is infinite. Suppose $h(n)$ is available exactly. Then the problem is to select a finite nonzero sequence $k(n)$ with approximately the same frequency response as that of $h(n)$. Intuitively, the larger values of $h(n)$ seem most important, so that if these values are selected out via a window $w(n)$, then the response of the result may be very close to the desired result.

This is the intuitive explanation of windowing from the time domain point of view.

From a frequency point of view, there is some frequency description of $h(n)$, viz., $H(s) = H'[z = \exp(j\omega)]$. Similarly for $w(n)$, $W(\omega) = W'[z = \exp(j\omega)]$. The frequency description of $k(n) = h(n) w(n)$ is proportional to $H(\omega) * W(\omega)$, where "*" indicates convolution in the complex domain. So that if the window $w(n)$ is to produce a FIR without substantially disturbing $H(\omega)$, then $W(\omega)$ ought to be nearly an impulse.

The problem is that $h(n)$ is not available. Only samples of $H(\omega)$ are available. However, a sequence $k(n)$ containing only a finite set of nonzero values is available. This sequence has the same frequency description as $h(n)$ at the sample points in the frequency domain. This $k(n)$ is the IDFT of those available samples.

Thus, the 5-T's method consists of the following steps. TRANSFORM the samples of the frequency description into the time domain. TRUNCATE the samples if a sample number limitation exists. TAILOR the samples via a window. TRANSFORM the resulting set back in the frequency domain. And TEST the results to see if it meets the specifications.

The windowing is a bit tricky. A discontinuous window will display the Gibbs phenomenon, so that it is a good idea to use a continuous one. The specific window used depends on the specific application. Note that no window is not possible; since there is a finite set of samples, the selecting of them implies at least a Fourier window. Before

transforming back into the frequency domain, it is necessary to augment the time sequence by a factor of 2 or 4 in order to get an adequate description of the results (not just at the original sampling points).

This appendix is meant as a short description of the 5-T's method and the reader is referred to the appropriate references for further details [6].

BIBLIOGRAPHY

1. A. Gelb and W. E. Vander Velde, *Multiple-Input Describing Functions and Nonlinear System Design*, MIT Press, Cambridge, Massachusetts, 1968.
2. B. C. Kuo, "A 'Z-Transform-Describing Function' for On-Off Type Sampled-Data Systems", *Proceedings IRE*, Vol. 48, January 1960, pp. 941-942.
3. D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley and Sons, Inc., New York, 1969.
4. R. C. Schramm and M. T. Scott, "Derivative Information Recovery by a Selective Integration Technique (DIRSIT)", Nike Zeus Data Report No. 35, Electro Mechanical Laboratories, White Sands Missile Range, New Mexico, May 1962.
5. R. C. Schramm and M. T. Scott, "Derivative Information Recovery by a Selective Integration Technique (DIRSIT)", Technical Memorandum No. 1040, D. A. Project 516-04-007, Electro Mechanical Laboratories, White Sands Missile Range, New Mexico, November 1962.
6. T. G. Stockham, Jr., unpublished class notes, 1972-1973.